A Multivariate Loss Ratio Approach for Systemic Risk Measurement and Allocation

Wei Wang* Huifu Xu†

September 11, 2025

Abstract

The primary challenges in systemic risk measurement involve determining an overall reserve level of risk capital and allocating it to different components based on their systemic relevance. In this paper, we introduce a multivariate loss ratio measure (MLRM), which is the minimum amount of capital to be injected into a financial system such that the ratio of the multivariate shortfall risk over the total capital to be injected falls within a specified degree of tolerance. The degree of the tolerance controls the balance between the expected systemic risk to be reduced and the amount of the capital to be injected. The MLRM recovers the well-known multivariate shortfall risk measure when the degree of tolerance is zero. Under some moderate conditions, we demonstrate that the MLRM is monotonic, continuous and convex, and that the optimal capital allocation based on the MLRM is unique. Moreover, we show that the risk capital allocation based on the MLRM optimizes the overall systemic performance (measured by a combination of the systemic multivariate shortfall risk and the cost of risk capital) with a minimum capital requirement. Furthermore, we analyze the sensitivity of the MLRM and the associated risk allocations with respect to variation of the degree of tolerance as well as underlying random data. Finally, we report numerical test results which highlight the efficiency and stability of the risk capital allocation under the MLRM.

Keywords: Risk management, multivariate loss ratio measure, risk tolerance degree, risk capital allocation, sensitivities.

1 Introduction

The global financial crisis of 2008-2009 and the COVID-19 pandemic have underscored the significance of addressing systemic risk and its potential implications (Rizwan et al., 2020). Systemic risk refers to the risk that a financial system is susceptible to severe instability or collapse due to the characteristics of the system itself (Chen et al., 2013; Feinstein et al., 2017; Ellis et al., 2022). Due to the tremendous cost of this type of risk, financial regulations, such as Basel III/IV for banking regulation and Solvency II

*Research Institute for Interdisciplinary Sciences, School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (wangwei1@sufe.edu.cn, math.wang21@gmail.com).

†Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong (hfxu@se.cuhk.edu.hk).

for insurance regulation, have been introduced to specifically address the systemic risk. Moreover, many empirical evidence shows that the interconnection among financial institutions has increased significantly in recent years, generating the risk of potential system-wide distress with major knock-on effects on the real economy (Chu et al., 2020; Veraart, 2020; Poledna et al., 2021). The ongoing concern about systemic risk has prompted intensive research on the design and implementation of tools for efficient systemic risk modeling, measurement, and management. Eisenberg and Noe (2001) seem the first to emphasize the need for a comprehensive approach to assess the financial sector's exposure to systemic risk. Over the past two decades, numerous studies on systemic risk have emerged for different purposes in the literature, see, e.g., Bisias et al. (2012) and Ellis et al. (2022) for the comprehensive literature review of systemic risk analytics and measurement respectively. In particular, various capital measures are introduced to identify the institutions' exposures under systemic events or to gauge the magnitudes of systemic risk of the whole risky financial system, see, e.g., Girardi and Ergün (2013); López-Espinosa et al. (2015); Adrian and Brunnermeier (2016); Acharya et al. (2017); Feinstein et al. (2017); Brownlees and Engle (2017); Armenti et al. (2018); Biagini et al. (2019); Doldi and Frittelli (2021); Cai et al. (2022); Wang et al. (2023) and references therein. Such tools are helpful for regulators to identify institutions that could significantly be affected by market shock and to assess the systemic risks of financial systems.

In this paper, we follow up this stream of research and study the systemic risk for financial systems with interconnected risky components by focusing on the following two main issues:

- The determination of an overall reserve level of risk capital such that the whole financial system can hedge the systemic risk under some degree of tolerance.
- The allocation of this overall amount of risk capital between the different risk components in a way that reflects the systemic risk of each one.

These two issues are of great importance in systemic risk management, as they both play a key role in ensuring the stability and resilience of interconnected financial systems. Specifically, the determination of risk capital meets essential requirements, such as the minimum capital requirement outlined in Basel III/IV. It serves as a financial cushion, strengthening the financial system against adverse events or stress scenarios. On the other hand, the proper allocation of the amount of risk capital enables the realization of the greatest risk reduction benefit. This allocation helps to decrease financial institutions' risk exposures, resulting in a significant reduction of the overall vulnerability of the system. In the existing literature, there are plenty of research on the similar focuses of the paper from different viewpoints and different approaches to various aspects of systemic risk, see, e.g., Chen et al. (2013); Feinstein et al. (2017); Armenti et al. (2018); Biagini et al. (2019); Brunnermeier and Cheridito (2019); Biagini et al. (2020); Doldi and Frittelli (2021, 2022); Wang et al. (2023) and references therein. To the best of our knowledge, the research falls into two distinct but sometimes overlapping categories, which are typically known as the "first aggregate-then-allocate type" and the "first allocate-then-aggregate type" of systemic risk measures, as discussed in, for example, Biagini et al. (2019). They are distinguished by their distinct methods of handling the issues of aggregation and allocation in the evaluation of the overall systemic risk associated to a risky financial system.

Chen et al. (2013) provide an axiomatic approach to the systemic risk measure of the first aggregate-then-allocate type on a prospect space with a finite number of scenarios. Specifically, they show that when the systemic risk measure satisfies the so-called preference consistency axiom and other natural axioms, there exists a unique aggregation function over cross-sectional profiles that is consistent across scenarios and a monetary risk measure such that the systemic risk measure is the composition of them. Moreover, they consider the allocation of the systemic risk across the components of the risky financial system based on the dual decomposition of the class of systemic risk measure. Their axiomatic approach is technically extended by Kromer et al. (2016) who work on a general measurable space and consider the case where the systemic risk measure are not necessary positively homogeneous, and by Hoffmann et al. (2016, 2018) who argue in favor of a conditional framework on general probability spaces for assessing systemic risk. Prominent systemic risk measures of this category include the systemic expected shortfall introduced by Acharya et al. (2017), the contagion value at risk (CoVaR) defined by Adrian and Brunnermeier (2016), and the SystRisk proposed by Brunnermeier and Cheridito (2019). These studies may provide regulators or policymakers with effective analytical tools for assessing systemic risks of financial systems. However, these systemic risk measures cannot apply to solve the second issue directly as the required capital is directly added into the whole system rather than the individual financial institutions. Consequently, it might be natural to address the aforementioned two issues simultaneously through the implementation of the first-allocate-then-aggregate type of systemic risk measures.

Armenti et al. (2018) first extend the well-known univariate shortfall risk measure introduced by Föllmer and Schied (2002) to multivariate case and propose the multivariate shortfall risk measure (MSRM) to measure the systemic risk of an interconnected system of risk components, where the required capital is first allocated among risk components and then aggregated such that the shortfall risk of the whole system is perfectly hedged. They emphasize existence, uniqueness, sensitivity and numerical applications of the associated deterministic risk allocation policy. Note that the systemic risk measure is calculated on the basis of the expected disutility of losses incurred over all scenarios leveraged by their respective objective probabilities rather than extreme scenarios as in stress tests, see, e.g., Sahin et al. (2020). In Biagini et al. (2019), they specify a general methodological framework for defining systemic risk measures via multidimensional acceptance sets and aggregation functions, and include the possibility of both deterministic and scenario-dependent allocation of required capital to secure the system. Interestingly, in their follow-up work, Biagini et al. (2020) prove a dual representation of the systemic risk measures for a class of additive aggregation functions, and the existence and uniqueness of the optimal allocation related to them. Moreover, they show that the optimiser in the dual formulation provides a natural risk allocation which is fair from the point of view of the individual financial institutions. Later, Doldi and Frittelli (2021); Doldi et al. (2023) extend their work further to the systemic risk measures for a class of more general aggregation functions. Recently, Wang et al. (2023) extend the work of Armenti et al. (2018) to allow scenario-dependent allocations of risk capital and propose a so-called optimal scenario-dependent MSRM to assess the systemic risk, which is less conservative than MSRM. Independently, a related concept in the context of set-valued systemic risk measures has been developed by a number of researchers such as Molchanov and Cascos (2016); Feinstein et al. (2017); Ararat et al. (2017) and references therein.

Note that the aforementioned systemic risk measures are based only on the system's overall stability and the decision maker's acceptance of multivariate shortfall risk of a risky financial system after capital injection, with explicit constraints on acceptability amounts of multivariate shortfall risk. However, in practice, the determination of risk capital for a risky financial system often involves a balance between safeguarding against systemic risk and managing the efficient use of risk capital. It is important to note that there are costs associated with providing risk capital, as highlighted in studies such as Laeven and

Goovaerts (2004). In this context, decision makers face the challenge of optimising risk protection while ensuring a cost-effective allocation of resources. In particular, allocating an excessive amount of capital for risk mitigation of the whole system can tie up resources that might otherwise be used for productive purposes. On the flip side, inadequate capital might render the system vulnerable to systemic shocks. Thus, from a decision maker's perspective, a typical trade-off in these settings revolves around finding the right balance between determining the appropriate amount of risk capital, the tolerable degree of multivariate shortfall risk, and the efficient use of risk capital. It is worth noting that the systemic risk measures reviewed do not take into account the efficient use of risk capital. As a result, the solutions they yield may be overly conservative and impractical.

In this paper, we aim to bridge this gap by introducing a novel framework which allows one to mitigate systemic risk of a risky financial system without committing excessive capital. The research is inspired by the recent work of Baron et al. (2023), where the authors introduce a new notion of loss ratio to safeguard capacity constraints in a medical resource allocation problem by limiting the relative expected shortage with respect to a variable capacity, and balance the risk of shortage and cost of resource deployment under data uncertainty. Specifically, we extend the modeling paradigm of Baron et al. (2023) to introduce a multivariate loss ratio measure (MLRM) to quantify systemic risk of a risky financial system by considering the trade-off between the multivariate shortfall risk and the capital to be injected as well as the decision maker's degree of risk tolerance. By incorporating these considerations, our methodology provides a comprehensive and effective strategy for measuring systemic risk. The main contributions of the paper can be summarized as follows.

First, we introduce a MLRM which is the minimum amount of capital to be injected into a financial system such that the ratio of the expected shortfall risk over the total capital to be injected falls within a specified degree of tolerance. The degree of the tolerance controls the balance between the expected systemic risk to be reduced and the amount of the capital to be injected. The MLRM recovers the well-known MSRM when the degree of tolerance is zero. It also effectively extends the loss ratio model of Baron et al. (2023) to multi-resource planning problems. The rationale behind the new measure is to allow the decision maker to mitigate systemic risk without committing excessive capital. Under some moderate conditions, we demonstrate that the MLRM is monotonic, continuous and convex. Some illustrative examples are provided to facilitate understanding of the proposed measure.

Second, we apply the proposed measure to the risk capital allocation problem and derive sufficient conditions which ensure existence and uniqueness of an optimal risk capital allocation policy. Moreover, we study the interplay between the multivariate loss function and the dependence structure of the components and highlight its relevance as an indicator of systemic risk. In particular, we provide an economic insight of the associated risk allocations by optimizing the overall systemic performance, which is described via the systemic multivariate shortfall risk plus the cost of risk capital, with minimum capital requirement.

Third, we investigate the sensitivity of the proposed measure and its associated risk capital allocations with respect to variation of risk tolerance level, random shock and perturbation of the uncertainty data. These analysis provides theoretical guidance for practical application of MLRM particularly in data-driven problems. To examine the performance of the MLRM and the associated risk capital allocation strategy, we have carried out a number of numerical experiments. The results show that the MLRM-based allocation strategies display similar properties as MSRM but with less capital requirement, stable with respect to small change of risk tolerance level and resilient with regard to uncertainty data perturbation.

The rest of the paper is organized as follows. In Section 2, we introduce the MLRM and discuss its key properties. In Section 3, we apply the proposed measure to the risk capital allocation problem. In Section 4, we analyze sensitivity of MLRM with respect to variation of various factors in the model. In Section 5, we report numerical test results of MLRM and associated capital risk allocations. Finally, we conclude with some remarks in Section 6. Some technical details are delegated to the appendices.

2 Multivariate loss ratio measure

Let Ω be a set of possible scenarios, and $(\Omega, \mathcal{F}, \mathbb{P})$ some abstract probability space with $\mathbb{P}$ a probability measure on $\mathcal{F}$, a σ-filed of events in Ω. The probability space is assumed to be atomless, so it is enough to support continuous distributions, see, e.g., Föllmer and Schied (2016). Let $L^0(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^n)$ (L_n^0 for short) be the space of all $\mathcal{F}$-measurable mappings $f : \Omega \mapsto \mathbb{R}^n$. For any $p \in [1, \infty)$, we denote by L_n^p the space of all measurable functions $f : \Omega \mapsto \mathbb{R}^n$ with $\|f\|_p < +\infty$, where the L^p-norm $\|\cdot\|_p$ is defined in terms of the probability measure $\mathbb{P}$ on Ω for each component. Let $\mathbf{X} = (X_1, \ldots, X_n)^{\mathrm{T}} \in L_n^0$ be a random vector representing a risky financial system with n components, where X_i stands for the risky position for i-th component; that is, positive values of X_i represents financial losses. For $\mathbf{X}, \mathbf{Y} \in L_n^0$, the inequality $\mathbf{X} \geq \mathbf{Y}$ is understood $\mathbf{X}(\omega) \geq \mathbf{Y}(\omega)$ component-wise for all $\omega \in \Omega$. For a probability measure Q on $(\Omega, \mathcal{F})$ we write $Q \ll \mathbb{P}$ for Q being absolutely continuous w.r.t. $\mathbb{P}$, and denote by $\frac{dQ}{d\mathbb{P}} \in L_1^1$ the corresponding Radon-Nikodym derivative. For a vector of probability measures $\boldsymbol{Q} = (Q_1, \ldots, Q_n)$, we write $\boldsymbol{Q} \ll \mathbb{P}$ for component-wise absolute continuity, i.e., $Q_i \ll \mathbb{P}$ for $i = 1, \ldots, n$. We set

$$\mathcal{Q} := \{\boldsymbol{Q} = (Q_1, \ldots, Q_n)^{\mathrm{T}} \text{ of probability measures on } (\Omega, \mathcal{F}), \text{s.t. } \boldsymbol{Q} \ll \mathbb{P}\}$$

and denote by $\frac{d\boldsymbol{Q}}{d\mathbb{P}}$ the vectors of Radon-Nikodym derivatives of $\boldsymbol{Q} \in \mathcal{Q}$.

To facilitate reading, we use the following notation throughout the paper. In $\mathbb{R}^n$, we denote $\mathbf{0} = (0, \ldots, 0)^{\mathrm{T}}$, $\mathbf{1} = (1, \ldots, 1)^{\mathrm{T}}$. Let $\boldsymbol{e}_k$ be the k-th unit vector and $B(\mathbf{0}, \Delta)$ the ball centered at $\mathbf{0}$ with radius $\Delta > 0$. By slight abuse of notation, we also use $\|\cdot\|_1$ to denote the l_1-norm for vectors in $\mathbb{R}^n$. Given an l_1-norm $\|\cdot\|_1$ in $\mathbb{R}^n$, that is, $\|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|$, let $d(\boldsymbol{x}, A) := \inf_{\boldsymbol{y} \in A} \|\boldsymbol{x} - \boldsymbol{y}\|_1$ be the distance from a point $\boldsymbol{x}$ to a set $A \subset \mathbb{R}^n$. For any two non-empty sets $A, B \subset \mathbb{R}^n$, we write $\mathbb{D}(A, B) := \sup_{\boldsymbol{x} \in A} d(\boldsymbol{x}, B)$ for the deviation of A from B and $\mathbb{H}(A, B) := \max\{\mathbb{D}(A, B), \mathbb{D}(B, A)\}$ for the Hausdorff distance between A and B. We employ bold notation to represent vectors, irrespective of whether they are deterministic, random, or vectors of probability measures.

In this paper, we propose a pragmatic approach to determine the overall reserve level of risk capital for a risky financial system, incorporating a degree of risk tolerance and efficient use of capital, while simultaneously allocating the risk capital to its various components. To this end, we introduce a new and flexible class of measures called MLRM, inspired by the recent work of Baron et al. (2023), by means of multivariate loss functions and a pre-specified safety or risk tolerance level.

2.1 Multivariate loss functions

To give a formal definition of the novel and flexible class of measures, we begin with a multivariate loss function l defined on $\mathbb{R}^n$. From an economic point of view, a multivariate loss function should have two basic properties: "the more losses, the riskier" and "diversification should not increase risk", see, Drapeau and Kupper (2013) for related discussions.

Definition 2.1 (Multivariate loss function) *A function $l : \mathbb{R}^n \mapsto \mathbb{R}$ is called a* multivariate loss function *if it is non-decreasing, non-constant along each direction $d \in \mathbb{R}^n_+$, convex, and lower semicontinuous over $\mathbb{R}^n$. In particular, l is called a* normalized multivariate loss function *if l also satisfies $l(\mathbf{0}) = 0$ and $l(\mathbf{1}) = n$.*

The terminology "loss function" is extracted from Föllmer and Schied (2002), where they introduce shortfall risk measure of univariate random loss. Armenti et al. (2018) consider multivariate loss function when they extend the univariate shortfall risk measure to multivariate shortfall risk measure, see also Wang et al. (2023). Here, we adopt the notion. Note also that the multivariate loss function defined as such coincides with the so-called "aggregation function" in Chen et al. (2013); Biagini et al. (2019), which aggregates the cross-sectional loss profiles of a risky financial system in a single scenario into a real number. Alternatively, it can be viewed as multivariate disutility functions, given its close connection to classical multivariate utility functions, see, e.g., Keeney and Raiffa (1993) and references therein. The multivariate loss function expresses a decision maker's risk preference over such loss profiles and convexity of l corresponds to certain risk aversion*. Throughout the paper, we use the terminology multivariate loss function and multivariate disutility function interchangeably. Normalization of multivariate loss function is used to specify the scale as a multivariate disutility function is unique up to positive scaling. Similar normalization conditions can be found in Chen et al. (2013); Armbruster and Delage (2015); Zhang et al. (2022).

In the economics and financial literature, particularly the areas of risk measurement and capital allocation, there have been many multivariate loss functions introduced for various purposes, see, e.g., Chen et al. (2013); Armenti et al. (2018); Doldi et al. (2023) and references therein. We list some of them here and in the Appendix A.1.

Example 2.1 (Armenti et al. (2018)) *Let $h : \mathbb{R} \mapsto \mathbb{R}$ be a one dimensional loss function satisfying Definition 2.1, such as*

$$h(x) = x_+ - \beta x_-,\ 0 \le \beta < 1, \quad h(x) = x + \frac{1}{2}(x_+)^2, \quad h(x) = e^x - 1,$$

where $x_+ := \max\{x, 0\}$ and $x_- = \max\{-x, 0\}$. Based on these, we obtain the following multivariate loss functions:

$$l_1(\boldsymbol{x}) = h\left(\sum_{i=1}^n x_i\right), \quad l_2(\boldsymbol{x}) = \sum_{i=1}^n h(x_i), \quad l_3(\boldsymbol{x}) = \alpha h\left(\sum_{i=1}^n x_i\right) + (1-\alpha)\sum_{i=1}^n h(x_i),\ 0 \le \alpha \le 1.$$

*Note that, in the literature, there are two distinct definitions of risk aversion in multivariate case. One is by Richard (1975) and the other is by Russell and Seo (1978). In this paper, we adopt the latter as it is a straightforward extension of the one in univariate case and is also closely related to the multivariate stochastic dominance. For the difference of these two definitions, we refer readers to Scarsini (1988).

The multivariate loss function l_1 is designed to measure the disutility loss on the aggregate risk suffered by institutions across the whole financial system whereas l_2 emphasizes the aggregated individual institution's disutility loss in the system. The later allows to adopt different loss functions for individual institutions. Furthermore, l_3 is a convex combination of l_1 and l_2 and the parameter α represents the relative importance of the corresponding disutility losses.

In practice, it might be difficult to determine a precise multivariate loss function for a risky financial system. However, we may adopt the techniques for eliciting multivariate utility function, see, e.g., Clemen and Reilly (2013); González-Ortega et al. (2018); Delage et al. (2022), and then use the transformation $l(\boldsymbol{x}) = -u(-\boldsymbol{x})$ based on the elicited multivariate utility function u to obtain the multivariate disutility function. Moreover, if the intrinsic structural dependence or interaction of the components in the risky financial system is known, we may obtain the multivariate loss function by its mechanism, see, e.g., Eisenberg and Noe (2001); Chen et al. (2013); Weber and Weske (2017). We also refer readers to Zhang et al. (2022); Wu et al. (2023) for the case that information on the true multivariate loss function/utility function in decision making problems is incomplete and subsequently a preference robust argument may be employed.

2.2 Multivariate shortfall risk

In this subsection, we introduce another crucial ingredient for defining the proposed multivariate loss ratio measure. To ensure clarity and readability, we first review the concepts of univariate shortfall risk and then introduce the multivariate shortfall risk. The idea of shortfall risk can be traced back to the work of Roy (1952), who initially explores asset allocation by minimizing the probability that an investor might not achieve their desired investment target, which is called the safety-first principle in the literature. In other words, shortfall risk consists of falling short of a specified target outcome. Subsequently, this idea has been further developed and applied in various fields, including finance, where it plays a fundamental role in understanding and managing risk in investment portfolios and financial systems, see, e.g., Fishburn (1977); Föllmer and Leukert (1999); Brown and Sim (2009); Brown et al. (2012); Cai et al. (2022); Long et al. (2023) and references therein. It seems that the formal definition of shortfall risk is originally introduced by Föllmer and Leukert (2000), which is defined as the expectation of the shortfall weighted by some univariant loss function. A typical example of shortfall risk is the lower partial moments which is well known in the economics literature, see, e.g., Bawa and Lindenberg (1977); Harlow and Rao (1989). In this paper, we extend the concept of univariate shortfall risk of Föllmer and Leukert (2000) to multivariate case.

Definition 2.2 (Multivariate shortfall risk) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a multivariate loss function and $\boldsymbol{\tau} \in \mathbb{R}^n$ be an allocation of the initial capital $z := \sum_{i=1}^n \tau_i$. The* multivariate shortfall risk *of $\boldsymbol{X} \in L_n^0$ is defined as the expectation $\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})]$ of $\boldsymbol{X} - \boldsymbol{\tau}$ weighted by the multivariate loss function l provided that the expectation is finite.*

Unlike the univariate shortfall risk, here we take into account the negative part, thereby extending the concept to better capture the risk in a multivariate context. When $n = 1$ and l vanishes on $(-\infty, 0]$, the definition recovers Föllmer and Leukert (2000, Definition 2.1) since $\mathbb{E}[l(X - \tau)] = \mathbb{E}[l((X - \tau)_+)]$. In this paper, we do not impose restriction of the domain of l to $\mathbb{R}_+$ to provide a more flexible definition of multivariate shortfall risk, see the same relaxation used in Föllmer and Schied (2002). Furthermore, this relaxation enables us to manage the magnitudes of the multivariate shortfall $(\boldsymbol{X}-\boldsymbol{\tau})_+$ and multivariate excess $(\boldsymbol{X}-\boldsymbol{\tau})_-$ above and below the "optimal target" allocation $\boldsymbol{\tau}$. This relaxation is important in risk management because it enables effective management of potential losses (shortfall) and gains (excess) in the financial system. Investors can tailor risk allocation strategies to the specific needs by adjusting trade-offs between these components. Note that the multivariate shortfall risk not only depends on the choice of multivariate shortfall risk and the initial capital, but also depends on the allocation. The allocation, represented by the vector $\boldsymbol{\tau}$, determines how the total initial capital is distributed among the individual components. As a result, different allocation schemes can lead to varying levels of multivariate shortfall risk, even when utilizing the same multivariate loss function and initial capital. The next proposition gathers the main properties of the multivariate shortfall risk.

Proposition 2.1 (Properties of multivariate shortfall risk) *Let $l:\mathbb{R}^n \mapsto \mathbb{R}$ be a multivariate loss function. For any $\boldsymbol{X} \in L_n^0$ and $\boldsymbol{\tau} \in \mathbb{R}^n$, if the expectation $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$ is finite, then the mapping $(\boldsymbol{X},\boldsymbol{\tau}) \mapsto \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$ is convex and lower semicontinuous, non-decreasing in $\boldsymbol{X}$, and non-increasing in $\boldsymbol{\tau}$.*

2.3 Multivariate loss ratio measure

We now turn to formally introduce the multivariate loss ratio measure for determining the amount of risk capital withheld by a risky financial system to overcome unexpected losses under some degree of conservatism, while simultaneously allocating the risk capital to its various components. The primary rational for the proposed measure is that investors are unwilling to commit the necessary capital for a perfect hedge or to cover unexpected losses. Instead, they prefer to withhold a smaller amount of capital to mitigate unexpected losses and are willing to accept some level of risk.

Definition 2.3 (Multivariate loss ratio measure (MLRM)) *Let $l:\mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function and $\delta \geq 0$ be a prespecified risk tolerance degree. The* multivariate loss ratio measure *of $\boldsymbol{X} \in L_n^0$ with risk tolerance degree δ is defined as*

$$\text{(MLRM)} \quad \rho_l^\delta(\boldsymbol{X}) := \inf_{\boldsymbol{\tau}\in\mathbb{R}^n} \quad \sum_{i=1}^n \tau_i \tag{2.1a}$$

$$\text{s.t.} \quad \mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{\tau})\right] \leq \delta \sum_{i=1}^n \tau_i, \tag{2.1b}$$

provided that the expectation is finite.

From the definition, we can see immediately that MLRM coincides with *multivariate shortfall risk measure* (MSRM) proposed by Armenti et al. (2018) when $\delta = 0$. Thus, we may view the former as an extension of the latter. To see the difference of the two quantities from finance perspective, we note that MSRM is the minimum cash to be injected, along an admissible allocation, into a risky financial system such that the multivariate shortfall risk is less than or equal to zero. In other words, MSRM is aimed at achieving a perfect hedge or complete coverage of expected losses in the sense that the overall multivariate shortfall risk is negative. From a practical point of view, the cost of achieving a perfect hedge could be high in most cases. Moreover, achieving a perfect hedge might reduce the opportunity of making a profit, together with the risk of incurring a loss. The MLRM model focuses on the case that $\delta > 0$ and addresses the issue by considering expected loss $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$ relative to the total amount of risk capital to be allocated $\sum_{i=1}^n \tau_i$ which captures the efficiency of capital allocation in the reduction of systemic risk. The replacement of the constraint on the absolute expected shortfall risk $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$ with the relative expected shortfall risk $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]/\sum_{i=1}^n \tau_i$ in (2.1b) allows the tolerance on the absolute expected shortfall risk to grow linearly in proportionate to the total amount of risk capital. The model switch from the absolute risk to relative risk is consistent with the practice where most investors' risk tolerance levels are usually dependent on the level of initial capital preparation $\sum_{i=1}^n \tau_i$ and higher level of initial capital requirement means a lower level of efficiency in risk management.

The MLRM can be interpreted as the minimum amount of capital to be injected such that the relative multivariate shortfall risk $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]/\sum_{i=1}^n \tau_i$ is bounded by a prescribed risk tolerance level δ. A smaller δ implies a greater degree of conservatism. Like $l(\cdot)$, the choice of δ is dependent on the decision maker's risk preference albeit the former is more on the DM's general attitude towards systemic risk whereas the latter is more specifically about the degree of the expected shortfall relative to the total capital (to be injected) that is acceptable by the DM. A more risk averse DM (corresponding to $l(\cdot)$ with greater curvature) may have a lower risk tolerance (small δ) or a higher risk tolerance (a larger δ) because the choice of $l(\cdot)$ is not directly related to the choice of δ. Moreover, the rhs of (2.1b) may be viewed as a linear disutility of the total capital and the optimum is attained when the expected disutility of the systemic risk ($\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$) is reduced to the level of the disutility of the capital deployment ($\delta\sum_{i=1}^n \tau_i$), see Theorem 5.1(ii). This is a departure from MSRM where the reduction of the expected disutility of the systemic risk ($\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$) is not related to the disutility of the capital deployment.

In practice, the total capital allocation is usually positive. However, purely from mathematical modelling perspective, we may allow $\sum_{i=1}^n \tau_i < 0$. In that case, constraint (2.1b) can be equivalently written as $-\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]/(-\sum_{i=1}^n \tau_i) \geq \delta$. If we interpret $l(\cdot)$ as a disutility function and $-l(-\cdot)$ as a utility function, then the inequality means that the ratio of the expected utility value of the remaining portfolio after withdrawal of a total amount of cash $-\sum_{i=1}^n \tau_i$ relative to the present monetary value of the cash must be greater or equal to δ.

Note that in Definition 2.3, we require l to be a normal multivariate loss function in order to ensure the well-definedness of $\rho_l^\delta(\boldsymbol{X})$. This is because multivariate disutility function l is unique up to scaling and for any positive constant $c \in \mathbb{R}$, one has $\rho_{cl}^{c\delta}(\boldsymbol{X}) = \rho_l^\delta(\boldsymbol{X})$ for any $\boldsymbol{X} \in L_n^0$ provided that $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$ is finite. Moreover, the extended new loss ratio measure may provide an alternative way to examine systemic risks in finance by taking into account the relative multivariate shortfall risk, rather than solely focusing on the absolute multivariate shortfall risk. It is crucial to note that the proposed MLRM differs substantially from the general framework of the systemic risk measure introduced in Feinstein et al. (2017) as our measure do not satisfy the translation invariance, although the latter is very general and subsumes a wide range of interesting examples, for instance, multivariate shortfall risk measure (Armenti et al., 2018) and CoVaR (Adrian and Brunnermeier, 2016).

Note also that when $n = 1$ and $l(t) = \max\{t, 0\}$, (2.1) coincides with the loss ratio model recently proposed by Baron et al. (2023), where X represents demand such as medical resources and τ is the allocation of the required resources. In that case, τ is restricted to take non-negative values. Here we allow τ_i to take negative values because in the context of capital allocation, it is possible to have negative allocations. Thus, the MLRM model may be regarded as a generalization of the loss ratio model to multivariate resource planning problem. Moreover, the loss function l allows one to capture a decision maker's risk disutility/risk preference.

Remark 2.1 *It might be interesting to consider the case when $\delta < 0$. In that case, we may rewrite (2.1) as*

$$\text{(MLRM-V1)} \quad \rho_l^\delta(\boldsymbol{X}) := \inf_{\boldsymbol{\tau}\in\mathbb{R}^n} \quad \sum_{i=1}^n \tau_i \tag{2.2a}$$

$$\text{s.t.} \quad \mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{\tau})\right] - \delta\sum_{i=1}^n \tau_i \le 0, \tag{2.2b}$$

where $-\delta$ may be interpreted as the unit cost of the risk capital. Laeven and Goovaerts (2004) seem to be the first to consider the cost of risk capital which is typically equal to the (opportunity) cost of capital charged by the shareholders. Wang et al. (2023) incorporate the cost into a scenario-dependent multivariate shortfall risk model for risk capital allocation but have not gone into theoretical analysis of the resulted shortfall risk measure. It is easy to observe that $\rho_l^\delta(\boldsymbol{X}) \ge \rho_l^0(\boldsymbol{X})$. The (MLRM-V1) model can be further extended to

$$\text{(MLRM-V2)} \quad \hat{\rho}_l^\delta(\boldsymbol{X}) := \inf_{\boldsymbol{\tau}\in\mathbb{R}^n} \quad \sum_{i=1}^n \tau_i \tag{2.3a}$$

$$\text{s.t.} \quad \mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{\tau})\right] \le C - (-\delta)\sum_{i=1}^n \tau_i, \tag{2.3b}$$

where C is a constant. In that case, constraint (2.3b) may be interpreted as the requirement of shortfall risk after allocation of risk capital being less or equal to C less $(-\delta)\sum_{i=1}^n \tau_i$. Alternatively, we can also interpret (2.3b) as the overall disutility value after allocation (shortfall loss plus the cost of risk capital), $\mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{\tau})\right] + (-\delta)\sum_{i=1}^n \tau_i$, does not exceed C. In the latter case, model (MLRM-V2) may be related to a multivariate version of optimized certainty equivalent (Ben-Tal and Teboulle, 2007). To see this clearly, consider the case that $l(x) = (\sum_{i=1}^n x_i)_+$ and $\delta \in (-1, 0)$. Then we have

$$\mathrm{CVaR}_{1+\delta}\left(\sum_{i=1}^n X_i\right) = \inf_{\sum_{i=1}^n \tau_i} \sum_{i=1}^n \tau_i + \frac{1}{1-(1+\delta)}\mathbb{E}\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \tau_i\right)_+\right]. \tag{2.4}$$

Let $\boldsymbol{\tau}$ be a feasible solution of (2.3). Then

$$(-\delta)\mathrm{CVaR}_{1+\delta}\left(\sum_{i=1}^n X_i\right) \le \mathbb{E}\left[\left(\sum_{i=1}^n X_i - \tau_i\right)_+\right] + (-\delta)\sum_{i=1}^n \tau_i \le C.$$

Since the optimal value of the minimization problem in (2.4) is attained at $\mathrm{VaR}_{1+\delta}(\sum_{i=1}^n X_i)$ and which is only a feasible solution of problem (2.3), then we conclude that

$$\hat{\rho}_l^\delta(\boldsymbol{X}) \le \mathrm{VaR}_{1+\delta}\left(\sum_{i=1}^n X_i\right) \le \mathrm{CVaR}_{1+\delta}\left(\sum_{i=1}^n X_i\right) \le -\frac{C}{\delta}.$$

It will be interesting to explore properties of multivariate loss ratio measures defined as in (MLRM-V1) and (MLRM-V2), but we leave it for future research as our focus in this paper is on MLRM.

Purely from optimization perspective, we note that problem (MLRM) is a convex program because the constraint function is convex and the objective function is linear. To ease the discussions, let

$$\mathcal{F}^{\delta}(\boldsymbol{X}) := \left\{ \boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^{n} \tau_i \right\} \tag{2.5}$$

be the feasible set of problem (2.1) and

$$\mathcal{F}^{\delta}_{=}(\boldsymbol{X}) := \left\{ \boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] = \delta \sum_{i=1}^{n} \tau_i \right\}. \tag{2.6}$$

Let $\mathcal{S}^{\delta}(\boldsymbol{X})$ denote the set of optimal solutions to problem (2.1). Then $\mathcal{S}^{\delta}(\boldsymbol{X})$ is a convex set and $\mathcal{S}^{\delta}(\boldsymbol{X}) \subset \mathcal{F}^{\delta}_{=}(\boldsymbol{X})$. Moreover, if the optimum is attained in the region

$$\Upsilon^{\delta}(\boldsymbol{X}) := \left\{ \boldsymbol{\tau} \in \mathbb{R}^n : \sum_{i=1}^{n} \tau_i \geq 0 \right\} \cap \mathcal{F}^{\delta}(\boldsymbol{X}), \tag{2.7}$$

then an increase of parameter δ will increase the size of $\Upsilon^{\delta}(\boldsymbol{X})$ and subsequently reduce the MLRM. On the other hand, if the optimal value is negative, then an increase of parameter δ will increase the MLRM. To facilitate understanding of the MLRM, we have include an example in the Appendix A.2 to illustrate how the MRLM may be worked out when the loss function takes a specific forms.

3 Natural domain of MLRM

Before proceeding to discuss the properties of the MLRM, we investigate the natural domain of the measure, that is, the class of random vectors $\boldsymbol{X}$ such that $\rho^{\delta}_{l}(\boldsymbol{X})$ is finite-valued as in Armenti et al. (2018) for the MSRM. For this purpose, we begin by recalling some basic notions and main results about the multivariate Orlicz space established in Armenti et al. (2018), see details in Appendix A.3. We also refer to Haezendonck and Goovaerts (1982); Cheridito and Li (2009) for the seminal work of Orlicz premiums and risk measures on Orlicz hearts, respectively.

Assumption 3.1 *There exist $a \in \mathbb{R}_+$ and $b \in \mathbb{R}$ such that the multivariate loss function $l : \mathbb{R}^n \mapsto \mathbb{R}$ satisfies $l(\boldsymbol{x}) \geq a \sum_{i=1}^{n} x_i + b$ for any $\boldsymbol{x} \in \mathbb{R}^n$.*

The assumption means that the epigraph of $l(\cdot)$ is supported or contained by the epigraph of the linear function $a \sum_{i=1}^{n} x_i + b$, which means that the left directional derivative of $l(\cdot)$ is lower bounded by a for any direction $\boldsymbol{d} \in \mathbb{R}^n_+$. Since the multivariate loss function is unique up to scaling, we can see that Assumption 3.1 is essentially equivalent to Armenti et al. (2018, Definition 2.1 (A3)). In their definition, they describe this assumption as a form of risk aversion, where the multivariate loss function puts more weight on high losses than a risk neutral evaluation. This assumption is less conservative, as it encompasses a wide range of multivariate loss functions. It is worth noting that the multivariate loss functions mentioned in Examples 2.1, A.1, A.2, and A.3 (except for l_2) all satisfy Assumption 3.1. Since l satisfies Assumption 3.1, then one has

$$\left\{\boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^{n} \tau_i \right\} \subseteq \left\{\boldsymbol{\tau} \in \mathbb{R}^n : a \sum_{i=1}^{n} \mathbb{E}[X_i] - a \sum_{i=1}^{n} \tau_i + b \leq \delta \sum_{i=1}^{n} \tau_i \right\}.$$

Consequently, by the definition of $\rho_l^\delta(\boldsymbol{X})$, one has

$$\rho_l^\delta(\boldsymbol{X}) \geq \frac{1}{a+\delta}\left(a \sum_{i=1}^{n} \mathbb{E}[X_i] + b\right). \tag{3.1}$$

Thus, $\rho_l^\delta(\boldsymbol{X}) > -\infty$ when $\boldsymbol{X} \in L_n^1$.

Proposition 3.1 (Natural domain of MLRM vs multivariate Orlicz space) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function satisfying Assumption 3.1 and $\delta > 0$ be a prespecified risk tolerance degree. Then the natural domain of the MLRM is the multivariate Orlicz space L_n^θ, where $\theta(\boldsymbol{x}) = l(|\boldsymbol{x}|)$ for $\boldsymbol{x} \in \mathbb{R}^n$.*

It is important to note that when $\delta = 0$, we cannot ensure that the natural domain of the MSRM introduced by Armenti et al. (2018) is the multivariate Orlicz space in our context unless 0 lies in the interior of the range of l. Thus, our model is more flexible than their model.

4 Main properties

Like the MSRM, the introduced MLRM enjoys a number of desirable properties, as stated in the next theorem.

Theorem 4.1 (Main properties of MLRM) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function satisfying Assumption 3.1 and $\delta > 0$ be a prespecified risk tolerance degree. Then the following assertions hold.*

(i) The MLRM is monotonic and convex on Orlicz space L_n^θ with $\theta(\boldsymbol{x}) = l(|\boldsymbol{x}|)$ for $\boldsymbol{x} \in \mathbb{R}^n$.

(ii) The MLRM admits the dual representation

$$\rho_l^\delta(\boldsymbol{X}) = \sup_{\boldsymbol{Q} \in \mathcal{Q}} \left\{ \sum_{i=1}^{n} \mathbb{E}_{Q_i}[X_i] - \alpha(\boldsymbol{Q}, \boldsymbol{X}) \right\},$$

where $\mathcal{Q} := \{\boldsymbol{Q} = (Q_1, \ldots, Q_n)^{\mathrm{T}} \text{ of probability measures on } (\Omega, \mathcal{F}), \text{s.t. } \boldsymbol{Q} \ll \mathbb{P}\}$,

$$\alpha(\boldsymbol{Q}, \boldsymbol{X}) := \inf_{\lambda > 0} \left\{ \lambda \mathbb{E}\left[l^*\left(\frac{1 - \lambda\delta}{\lambda} \frac{d\boldsymbol{Q}}{d\mathbb{P}} \right) \right] + \lambda\delta \sum_{i=1}^{n} \mathbb{E}_{Q_i}[X_i] \right\},$$

and $l^*(\cdot)$ is the convex conjugate of l.

(iii) Let $\boldsymbol{X} \in L_n^\theta$ be fixed. Assume that there exists a positive constant Δ such that $\mathcal{S}^\delta(\boldsymbol{Y}) \cap B(\mathbf{0}, \Delta) \neq \emptyset$ for all $\boldsymbol{Y}$ in a neighborhood of $\boldsymbol{X}$ equipped with Luxemburg norm $\|\cdot\|_\theta$. Then $\rho_l^\delta(\cdot)$ is continuous at $\boldsymbol{X}$.

Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function. For any $\gamma > 0$, we consider a one parameterized multivariate loss function,

$$l_\gamma(\boldsymbol{x}) := \frac{l(\gamma \boldsymbol{x})}{\gamma}, \ \forall\, \boldsymbol{x} \in \mathbb{R}^n. \tag{4.1}$$

Then for any $\delta \geq 0$, the corresponding MLRM is

$$\begin{aligned} \rho_\gamma(\boldsymbol{X}) := \rho_{l_\gamma}^\delta(\boldsymbol{X}) &= \inf\left\{\sum_{i=1}^n \tau_i : \mathbb{E}[l_\gamma(\boldsymbol{X} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^n \tau_i\right\} \\ &= \inf\left\{\sum_{i=1}^n \tau_i : \mathbb{E}[l(\gamma\boldsymbol{X} - \gamma\boldsymbol{\tau})] \leq \delta \sum_{i=1}^n \gamma\tau_i\right\} = \frac{1}{\gamma}\rho_l^\delta(\gamma\boldsymbol{X}). \end{aligned} \tag{4.2}$$

The next proposition states that $\rho_\gamma(\boldsymbol{X})$ is a nondecreasing function of γ and the MLRM possesses the superhomogeneity.

Proposition 4.1 (Super-homogeneity of MLRM) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function satisfying Assumption 3.1 and $\delta \geq 0$ be a specified risk tolerance level. Then for any $\boldsymbol{X} \in L_n^\theta$, $\rho_l^\delta(\boldsymbol{X})$ is superhomogeneous, i.e.,*

$$\rho_l^\delta(\gamma\boldsymbol{X}) \geq \gamma\rho_l^\delta(\boldsymbol{X}), \ \forall\, \gamma > 1, \quad \text{and} \quad \rho_l^\delta(\gamma\boldsymbol{X}) \leq \gamma\rho_l^\delta(\boldsymbol{X}), \ \forall\, 0 \leq \gamma \leq 1.$$

Moreover, the risk-to-exposure ratio $r_{\mathbf{X}} : \gamma \mapsto \rho_l^\delta(\gamma X)/\gamma$ is a nondecreasing function of γ on $(0, \infty)$.

The superhomogeneity of the MLRM corresponds to the star-shapedness of a risk measure considered in Castagnoli et al. (2022). Unfortunately the MLRM does not satisfy the translation invariance and normalization property since $\rho_l^\delta(\mathbf{0})$ may not equal to 0, see the counter example in Example A.3 with $l = l_1$.

4.1 MLRM preserves multivariate stochastic ordering

Multiattribute decision making under uncertainty assumes the ability of ranking random vectors. In this section, we show that the proposed MLRM is consistent with some widely-discussed multivariate stochastic ordering in the literature. Specifically, we follow the works of Rüschendorf (2004) and Armenti et al. (2018) to show that the proposed MLRM is monotone with regard to supermodularity, directional convexity, and upper orthant stochastic ordering as MSRM.

Definition 4.1 *For a function $f : \mathbb{R}^n \mapsto \mathbb{R}$, define the difference operator*

$$\Delta_{k,\boldsymbol{y}} f(\boldsymbol{x}) = f(x_1, \ldots, x_k + y_k, \ldots, x_n) - f(\boldsymbol{x}), \ \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n, \ k \in \{1, \ldots, n\}.$$

We say that a continuous function $f : \mathbb{R}^n \mapsto \mathbb{R}$ is

- supermodular *if* $\Delta_{k,\boldsymbol{y}}\Delta_{l,\boldsymbol{y}}f(\boldsymbol{x}) \geq 0$ *for every* $1 \leq k < l \leq n$*,* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $\boldsymbol{y} \geq 0$*;*
- directionally convex *if* $\Delta_{k,\boldsymbol{y}}\Delta_{l,\boldsymbol{y}}f(\boldsymbol{x}) \geq 0$ *for every* $1 \leq k \leq l \leq n$*,* $\boldsymbol{x} \in \mathbb{R}^n$ *and* $\boldsymbol{y} \geq 0$*;*
- Δ-monotone *if* $\Delta_{i_1,\boldsymbol{y}} \cdots \Delta_{i_k,\boldsymbol{y}}f(\boldsymbol{x}) \geq 0$ *for every* $\{i_1,\ldots,i_k\} \subseteq \{1,\ldots,n\}$ *with* $i_1 < \cdots < i_k$*,* $\boldsymbol{x} \in \mathbb{R}^n$*, and* $\boldsymbol{y} \geq 0$*.*

We denote by $\succeq^{sm}, \succeq^{dc}, \succeq^{uo}$ the induced integral stochastic orders on the class of probability measures (on $\mathbb{R}^n$) given by the respective class of functions. We refer to Rüschendorf (2004) for a discussion of these orders in terms of dependence risk. Note that $\succeq^{uo}$ is equivalent to the simple upper orthant order defined for $\boldsymbol{X} \succeq^{uo} \boldsymbol{Y}$ if and only if $\mathbb{P}[\boldsymbol{X} \geq \boldsymbol{x}] \geq \mathbb{P}[\boldsymbol{Y} \geq \boldsymbol{x}]$ for any $\boldsymbol{x} \in \mathbb{R}^n$.

Proposition 4.2 *The multivariate loss ratio measure is monotone with respect to* $\succeq^{sm}, \succeq^{dc}$*, or* $\succeq^{uo}$ *whenever* l *is supermodular, directionally convex, or* Δ*-monotone, respectively.*

Note that any multivariate loss function satisfying Assumption 3.1 is directionally convex and therefore supermodular. They are Δ-monotone if $n = 2$.

5 Risk capital allocation

In Section 2, we have established that the infimum over all admissible allocations $\boldsymbol{\tau} \in \mathbb{R}^n$, which is used to define the systemic risk $\rho_l^\delta(\cdot)$, is indeed a real-valued under some mild conditions and has several desirable properties. Beyond the consideration of the overall liquidity reserve, the allocation of this amount of risk capital among various risk components is significant importance in the context of systemic risk. Hence, in this section, we aim to tackle the following questions: (a) The existence of a risk allocation; (b) The uniqueness of a risk allocation; (c) The impact of the interdependence structure. The first question is important in some applications such as the default fund contribution of each member of a clearinghouse or the allocation of the capital among the different business lines of a bank. As for the second question, non-uniqueness can become an issue when this allocation is a regulatory cost for the different members or desks. If no additional clear rule is provided, the members would then face arbitrarinesses as for their contributions for the same overall risk. As for the last question, systemic risk should reflect the level of the dependence of the system. For instance, highly correlated losses, while having the same marginal risk, should result in a higher systemic risk and different optimal allocations. These questions are important in understanding and managing systemic risk, ensuring fairness and clarity in risk capital allocation, and aligning risk assessment with the underlying interconnections within financial systems.

A *risk capital allocation* is an acceptable monetary allocation $\boldsymbol{\tau} \in \mathbb{R}^n$ such that $\rho_l^\delta(\boldsymbol{X}) = \sum_{i=1}^n \tau_i$. By definition, if a risk capital allocation exists, then the full allocation property automatically holds. In contrast to the univariate case, the existence and uniqueness are no longer straightforward in the multivariate case. The next example provide a counter example.

Example 5.1 (Non-attainability of MLRM) *Consider the normalized multivariate loss function* $l(x_1, x_2) = 2(x_1 - 1)_+ + 2x_2$*. It is easy to observe that* $l(x_1, x_2) \geq 2(x_1 + x_2) - 2$ *and consequently* l *satisfies Assumption 3.1. Thus, for any* $\boldsymbol{X} \in L_2^\theta$*,* $\rho_l^\delta(\boldsymbol{X}) \in \mathbb{R}$*. For brevity, let* $\boldsymbol{X} = (X_1, X_2)^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$ *be a bivariate* *normal vector with*

$$\boldsymbol{\mu} = (0,0)^{\mathrm{T}} \text{ and } \Sigma = I := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then one has

$$\begin{aligned} \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] &= 2\mathbb{E}[(X_1 - \tau_1 - 1)_+] + 2\mathbb{E}[(X_2 - \tau_2)] \\ &= 2\int_{\tau_1+1}^{\infty} (x_1 - \tau_1 - 1)\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) dx_1 - 2\tau_2 \\ &= \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\tau_1+1)^2\right) + 2(-\tau_1 - 1)\left(1 - \Phi\left(\tau_1 + 1\right)\right) - 2\tau_2 \\ &= \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\tau_1+1)^2\right) + 2(\tau_1 + 1)\Phi(\tau_1 + 1) - 2(\tau_1 + \tau_2) - 2, \end{aligned}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. Consequently, one has

$$\begin{aligned} \rho_l^{\delta}(\boldsymbol{X}) &= \inf\left\{\tau_1 + \tau_2 : \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\tau_1+1)^2\right) + 2(\tau_1+1)\Phi(\tau_1+1) - 2 \le (2+\delta)(\tau_1+\tau_2)\right\} \\ &= -\frac{2}{2+\delta}. \end{aligned} \tag{5.1}$$

Define function $\psi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) + x\Phi(x)$. It is easy to verify that $\psi(x)$ is non-decreasing ($\psi'(x) = \Phi(x) \ge 0$) and $\lim_{x\to-\infty}\psi(x) = 0$. Then we can rewrite the constraint of the minimization problem in (5.1) as

$$-\frac{2}{2+\delta} \le \tau_1 + \tau_2 - \frac{\psi(\tau_1+1)}{2+\delta},$$

which means that the minimum is attained only when $\psi(\tau_1 + 1) \to 0$ or equivalently $\tau_1 \to -\infty$. This shows that the infimum is not attained at any finite τ_1, see figure A.4 for the graphic interpretation.

To address the non-attainability issue, we introduce conditions which ensures the existence and uniqueness of a risk allocation as well as the characterization of the risk capital allocation. The next theorem states this.

Assumption 5.1 *Let $\boldsymbol{X} \in L_n^{\theta}$ and $\mathcal{F}^{\delta}(\boldsymbol{X})$ be defined as in (2.5). Assume*

$$0^+\mathcal{F}^{\delta}(\boldsymbol{X}) \cap \mathcal{Z} = \{\mathbf{0}\}, \tag{5.2}$$

where $0^+\mathcal{F}^{\delta}(\boldsymbol{X})$ is the recession cone of $\mathcal{F}^{\delta}(\boldsymbol{X})$ as defined in Appendix A.5 and $\mathcal{Z} := \{\boldsymbol{\tau} \in \mathbb{R}^n : \sum_{i=1}^n \tau_i = 0\}$ is the set of zero-sum allocations.

Condition (5.2) is standard to secure the boundedness of the set of optimal solutions. This is evident from Examples A.3 and 5.1. The condition is slightly different from Armenti et al. (2018), where the authors assume $0^+l \cap \mathcal{Z} = \{\mathbf{0}\}$, where 0^+l denotes the recession cone of the function l (see Appendix A.5 for the definition). However, the key assumption remains the same as in Armenti et al. (2018).

Theorem 5.1 (Attainability of MLRM and uniqueness) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function satisfying Assumption 3.1 and $\delta > 0$ be a prespecified risk tolerance degree. Then* *the following assertions hold.*

(i) *For every $\boldsymbol{X} \in L_n^\theta$, if Assumption 5.1 is satisfied, then the minimum in problem (2.1) is attained in the set $\mathcal{F}^\delta_=(\boldsymbol{X})$ as defined in (2.6) and the set of optimal solutions is bounded and convex.*

(ii) *Any optimal solution $\boldsymbol{\tau}^* \in \mathbb{R}^n$ satisfies the following equations*

$$(1-\lambda^*\delta)\cdot \mathbf{1} \in \lambda^*\mathbb{E}[\partial l(\boldsymbol{X}-\boldsymbol{\tau}^*)] \quad \text{and} \quad \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau}^*)] = \delta\sum_{i=1}^n \tau_i^*, \tag{5.3}$$

where $\lambda^* > 0$ is the Lagrange multiplier and $\partial l(\boldsymbol{x})$ is the convex subdifferential of l at $\boldsymbol{x}$, see, e.g., Rockafellar (1997).

(iii) *If $l(\boldsymbol{x}+\cdot)$ is strictly convex along zero-sum allocations for every $\boldsymbol{x}$, then the optimal solution to problem (2.1) is unique.*

It is important to note that when $0^+\mathcal{F}^\delta(\boldsymbol{X}) \cap \mathcal{Z} \neq \{\mathbf{0}\}$, the problem (2.1) may be not attained, as illustrated in Example 5.1. Moreover, when $0^+\mathcal{F}^\delta(\boldsymbol{X}) \cap \mathcal{Z} \neq \{\mathbf{0}\}$, there are cases that the optimal solution set to problem (2.1) is attained but unbounded, as demonstrated in Example A.3. Finally, it is easy to verify that the risk capital allocation based on MLRM does not satisfy the cash additivity in general. However, it enjoys positive homogeneity when $l(\cdot)$ is positively homogeneous. We refer interested readers to Balog et al. (2017) and Guo et al. (2021) for a review and comparison of the main properties and methods of risk capital allocation.

Remark 5.1 *In general, the positivity of the risk allocation is not required. However, if positivity or any other convex constraint is imposed, for instance, by regulators, it can easily be embedded into our setup. In the case of positivity, this would modify the definition of the MLRM into*

$$\inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\left\{\sum_{i=1}^n \tau_i : \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})] \leq \delta\sum_{i=1}^n \tau_i \text{ and } \boldsymbol{\tau} \geq 0\right\}. \tag{5.4}$$

Under this case, the attainment to problem (5.4) is natural if the multivariate loss function l satisfies Assumption 3.1.

We now provide an example to illustrate the application of the MLRM in risk capital allocation when the multivariate loss function is the sum of exponential functions.

Example 5.2 (Illustration of uniqueness) *Consider normalized loss function*

$$l(x_1,\ldots,x_n) = \frac{1}{k}\left[\sum_{i=1}^n \exp\left(\alpha_i x_i\right) - n\right], \tag{5.5}$$

where $\alpha_i > 0$ for $i = 1,\ldots,n$ and $k = \sum_{i=1}^n \exp(\alpha_i)/n - 1$. The function satisfies Assumption 3.1, and is strictly convex along zero-sum allocations for every $\boldsymbol{x}$. By Theorem 5.1, the optimal allocation exists and is unique. In what follows, we derive the optimal allocation. Let $\boldsymbol{X} \in L_n^\theta$. We calculate the MLRM

$$\rho_l^\delta(\boldsymbol{X}) = \inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\left\{\sum_{i=1}^n \tau_i : \frac{1}{k}\mathbb{E}\left[\sum_{i=1}^n \exp\left(\alpha_i(X_i-\tau_i)\right) - n\right] \leq \delta\sum_{i=1}^n \tau_i\right\}. \tag{5.6}$$

By Theorem 5.1, we can write down the KKT conditions for the problem:

$$\sum_{i=1}^{n}\frac{\mathbb{E}[\exp(\alpha_i X_i)]}{\exp(\alpha_i\tau_i)}=k\delta\sum_{i=1}^{n}\tau_i+n,\ \text{and}\ \frac{1-\lambda\delta}{\lambda}\frac{k}{\alpha_i}=\frac{\mathbb{E}[\exp(\alpha_i X_i)]}{\exp(\alpha_i\tau_i)},\ i=1,\dots,n;$$

which yields

$$k\frac{1-\lambda\delta}{\lambda}\sum_{i=1}^{n}\frac{1}{\alpha_i}=k\delta\sum_{i=1}^{n}\tau_i+n \tag{5.7}$$

and

$$\tau_i=\frac{1}{\alpha_i}\ln\mathbb{E}\left[\exp(\alpha_i X_i)\right]-\frac{1}{\alpha_i}\ln\left(\frac{1-\lambda\delta}{\lambda}\right)+\frac{1}{\alpha_i}\ln\left(\frac{\alpha_i}{k}\right),\ i=1,\dots,n. \tag{5.8}$$

Thus

$$\frac{1-\lambda\delta}{\lambda}\sum_{i=1}^{n}\frac{1}{\alpha_i}=\delta\sum_{i=1}^{n}\frac{1}{\alpha_i}\ln\mathbb{E}\left[\exp(\alpha_i X_i)\right]-\delta\ln\left(\frac{1-\lambda\delta}{\lambda}\right)\sum_{i=1}^{n}\frac{1}{\alpha_i}+\delta\sum_{i=1}^{n}\frac{1}{\alpha_i}\ln\left(\frac{\alpha_i}{k}\right)+\frac{n}{k},$$

that is

$$\frac{1-\lambda\delta}{\lambda}+\delta\ln\left(\frac{1-\lambda\delta}{\lambda}\right)=\frac{1}{\sum_{i=1}^{n}\frac{1}{\alpha_i}}\left(\delta\sum_{i=1}^{n}\frac{1}{\alpha_i}\ln\mathbb{E}\left[\exp(\alpha_i X_i)\right]+\delta\sum_{i=1}^{n}\frac{1}{\alpha_i}\ln\left(\frac{\alpha_i}{k}\right)+\frac{n}{k}\right). \tag{5.9}$$

Since $\boldsymbol{X}\in L_n^{\theta}$, then the rhs of (5.9) is a constant in $\mathbb{R}$ denoted by C. Let $t=\frac{1-\lambda\delta}{\lambda}>0$. Then we can write the equation succinctly as $t+\delta\ln(t)=C$. Since the function $t\mapsto t+\delta\ln(t)$ is strictly increasing and unbounded from below and above, then equation has a unique solution, denoted by t^*. Consequently we have $\lambda^*=\frac{1}{\delta+t^*}$ and obtain via (5.7)-(5.8)

$$\rho_l^{\delta}(\boldsymbol{X})=\frac{t^*}{\delta}\sum_{i=1}^{n}\frac{1}{\alpha_i}-\frac{n}{k\delta},\ \text{and}\ \tau_i^*=\frac{1}{\alpha_i}\ln\mathbb{E}\left[\exp(\alpha_i X_i)\right]-\frac{1}{\alpha_i}\ln(t^*)+\frac{1}{\alpha_i}\ln\left(\frac{\alpha_i}{k}\right),\ i=1,\dots,n.$$

In the next example, we show the impact of the interdependence structure of data on the MLRM.

Example 5.3 *Let $\boldsymbol{X}=(X_1,X_2)^{\mathrm{T}}\sim\mathcal{N}(\mathbf{0},\Sigma)$ be a bivariate normal vector with covariance matrix $\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}$. Consider the normalized simplified counterpart of multivariate loss function in Example A.2, i.e.,*

$$l(x_1,x_2)=\frac{1}{\beta}\left(\frac{1}{2}\exp(2x_1)+\frac{1}{2}\exp(2x_2)+\exp(x_1+x_2)-2\right),$$

where $\beta=e^2-1>0$, we calculate the multivariate loss ratio measure

$$\rho_l^{\delta}(\boldsymbol{X})=\inf_{\boldsymbol{\tau}\in\mathbb{R}^2}\left\{\tau_1+\tau_2:\frac{1}{2\beta}\frac{\mathbb{E}[\exp(2X_1)]}{\exp(2\tau_1)}+\frac{1}{2\beta}\frac{\mathbb{E}[\exp(2X_2)]}{\exp(2\tau_2)}+\frac{1}{\beta}\frac{\mathbb{E}[\exp(X_1+X_2)]}{\exp(\tau_1+\tau_2)}-\frac{2}{\beta}\leq\delta(\tau_1+\tau_2)\right\}.$$

By solving the corresponding KKT system defined in Theorem 5.1, we obtain $\tau_2 - \tau_1 = \sigma_2^2 - \sigma_1^2$ *and*

$$\exp\left(\sigma_1^2+\sigma_2^2\right)+\exp\left(\frac{1}{2}(\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)\right)=(\beta\delta(\tau_1+\tau_2)+2)\exp\left(\tau_1+\tau_2\right). \tag{5.10}$$

In the case that $\sigma_1 \neq 0$ *or* $\sigma_2 \neq 0$*, the lhs of (5.10) is a constant, denoted by* $C > 2$*. Let* $t = \tau_1+\tau_2$*. We can rewrite equation (5.10) succinctly as* $(\beta\delta t+2)\exp(t) = C$*. Since the function* $t \mapsto \psi(t) := (\beta\delta t+2)\exp(t)$ *is strictly increasing and unbounded from above when* $t \geq 0$*, and* $\psi(0) = 2$*, then the equation has a unique solution, denoted by* t^**. Moreover, since* $C > 2$*, then* $t^* > 0$ *for all* δ*. In this case, the optimal allocation is* $(\tau_1^*, \tau_2^*) = \left(\frac{1}{2}t^* + \frac{1}{2}(\sigma_1^2 - \sigma_2^2), \frac{1}{2}t^* + \frac{1}{2}(\sigma_2^2 - \sigma_1^2)\right) \in \Upsilon(\boldsymbol{X})$*, where* $\Upsilon(\cdot)$ *is defined as in (2.7). Figures*

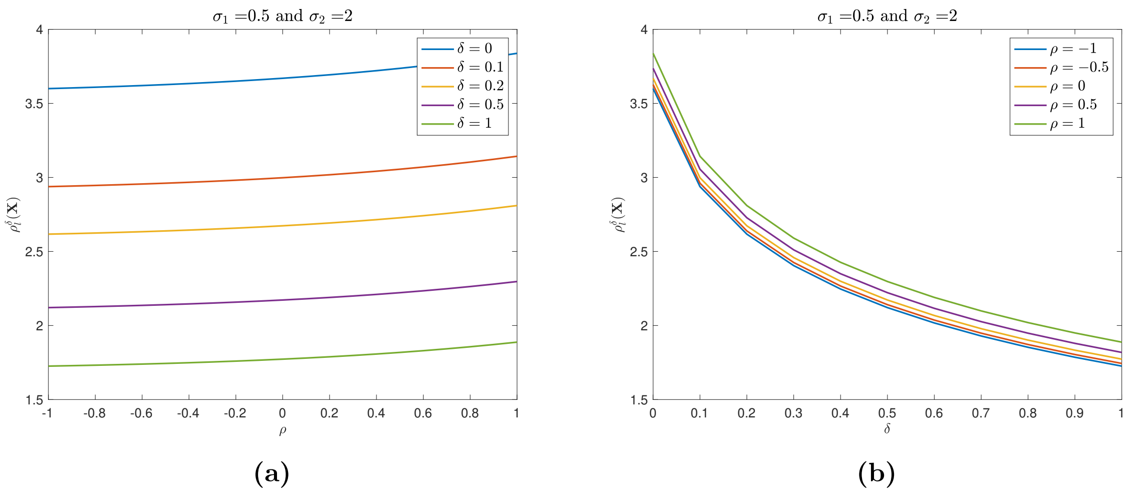


Figure 1: The value of MLRM as a function of **(a)** correlation parameter ρ and **(b)** tolerance level δ under different choice of parameters.

1(a) and (b) depict $\rho_l^\delta(\boldsymbol{X})$ *as a function of the correlation parameter* ρ *and the tolerance level* δ*. From Figure 1(a), we can see that* $\rho_l^\delta(\boldsymbol{X})$ *increases with respect to increase of the correlation coefficient* ρ *for fixed* δ*. This is because as* ρ *increases,* C *also increases and subsequently* t^* *increases. In contrast, Figure 1(b) shows the* $\rho_l^\delta(\boldsymbol{X})$ *decreases as* δ *increases. This is because* $(\tau_1^*, \tau_2^*) \in \Upsilon(\boldsymbol{X})$ *and as* δ *increases, the size of* $\Upsilon(\boldsymbol{X})$ *increases, see our earlier discussions immediately after (2.7). Moreover, it seems that the impact of the tolerance level on MLRM is more significant than the impact of correlation* ρ*.*

Before ending this section, we provide some economic insights of the induced optimal risk capital allocation strategy based on MLRM.

Theorem 5.2 (MLRM vs overall systemic performance) *Let* $(\boldsymbol{\tau}^*, \lambda^*)$ *be a saddle point of the Lagrangian function* $L(\boldsymbol{\tau}, \lambda; \boldsymbol{X})$ *of MLRM problem (2.1) with* $\lambda^* > 0$*. Then the following assertions hold:*

(i) $\boldsymbol{\tau}^*$ *is an optimal solution to the following problem*

$$\inf_{\boldsymbol{a}\in\mathbb{R}^n} \quad \mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{a})\right]+\frac{1}{\lambda^*}\sum_{i=1}^n a_i \tag{5.11a}$$

$$\text{s.t.} \quad \sum_{i=1}^n a_i \geq \rho_l^\delta(\boldsymbol{X}), \tag{5.11b}$$

and δ is the corresponding Lagrange multiplier.

(ii) Let $\psi(\delta) := \rho_l^\delta(\boldsymbol{X})$. Then

$$\psi'(\delta) = \inf_{\boldsymbol{\tau}^* \in S^\delta(\boldsymbol{X})} \sup_{\lambda^* \in \Lambda^\delta(\boldsymbol{X})} -\lambda^* e^{\mathrm{T}} \boldsymbol{\tau}^*, \tag{5.12}$$

where $S^\delta(\boldsymbol{X})$ and $\Lambda^\delta(\boldsymbol{X})$ denote respectively the set of optimal solutions and the set of Lagrange multipliers.

Let $\mathcal{T}(\boldsymbol{a}) := \mathbb{E}\left[l(\boldsymbol{X} - \boldsymbol{a})\right] + \frac{1}{\lambda^*} \sum_{i=1}^n a_i$. Then we may interpret $\frac{1}{\lambda^*}$ as the unit cost of the risk capital and consequently $\mathcal{T}(\boldsymbol{a})$ as overall systemic performance, which is described via the systemic multivariate shortfall risk plus the cost of risk capital, whereas the constraint of (5.11b) is the minimum capital requirement. From the first equation of (B.2), we can see that

$$\delta \cdot \mathbf{1} \in \lambda^* \mathbb{E}[\partial l(\boldsymbol{X} - \boldsymbol{\tau}^*)] + \frac{1}{\lambda^*} \mathbf{1} = \partial_{\boldsymbol{a}} \mathcal{T}(\boldsymbol{\tau}^*)$$

and hence we can interpret δ as the marginal change of $\mathcal{T}(\boldsymbol{a})$ with respect to the variation of a_i for $i = 1, \ldots, n$. Moreover, from (5.12), we can see that $\psi'(\delta) \le 0$ when $\rho_l^\delta(\boldsymbol{X}) \ge 0$ and $\psi'(\delta) \ge 0$ when $\rho_l^\delta(\boldsymbol{X}) < 0$. This means $\psi(\delta)$ attains maximum at $\delta = 0$.

6 Systemic sensitivity of the MLRM and the related allocation

Considering the systemic sensitivity of a new measure is important to risk management and decision making. It provides valuable insights into how variations in input parameters or data uncertainty affect the measure, and offers a deeper understanding of the model's behaviour. In this section, we discuss the sensitivity of the introduced MLRM and its induced risk allocation strategy with respect to perturbations of the risk tolerance level δ and the risky financial system $\boldsymbol{X}$, as well as the resilence against data perturbation.

6.1 Impact of risk tolerance level

We begin by analyzing the impact of perturbation of the risk tolerance level δ on the MLRM and the associated allocation strategy. The next theorem addresses this. Recall that $\mathcal{S}^\delta(\boldsymbol{X})$ denotes the set of optimal solutions to problem (2.1).

Theorem 6.1 (Stability of MLRM w.r.t. δ) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function and $\delta > 0$ be a prespecified risk tolerance level. Let $\boldsymbol{X} \in L_n^\theta$. Assume that there exists a positive constant Δ such that $\mathcal{S}^{\delta'}(\boldsymbol{X}) \cap B(\mathbf{0}, \Delta) \neq \emptyset$ for all positive δ' close to δ. Then the following assertions hold.*

(i) $\rho_l^{\delta'}(\boldsymbol{X})$ is Lipschitz continuous w.r.t. δ', i.e.,

$$\left|\rho_l^{\delta'}(\boldsymbol{X}) - \rho_l^\delta(\boldsymbol{X})\right| \le \max\left\{\frac{1}{\delta} \max\{0, (\delta' - \delta)\rho_l^{\delta'}(\boldsymbol{X})\}, \frac{1}{\delta'} \max\{0, (\delta - \delta')\rho_l^\delta(\boldsymbol{X})\}\right\} \tag{6.1}$$

for all δ' close to δ.

(ii) Assume in addition: (a) Assumption 5.1 holds, (b) $l(\cdot)$ is strictly convex. Define the growth function

$$\Lambda(\eta) := \min_{\boldsymbol{\tau}} \left\{ \sum_{i=1}^{n} \tau_i - \sum_{i=1}^{n} \tau_i^{\delta} : \|\boldsymbol{\tau} - \boldsymbol{\tau}^{\delta}\|_1 \geq \eta, \ \forall\, \boldsymbol{\tau} \in B(\mathbf{0}, \Delta) \right\},$$

where $\boldsymbol{\tau}^{\delta}$ is the unique optimal solution to the problem (2.1) with tolerance degree δ, and $\Lambda^{-1}(\nu) := \sup\{\eta : \Lambda(\eta) \leq \nu\}$. Then

$$\left\| \boldsymbol{\tau}^{\delta} - \boldsymbol{\tau}^{\delta'} \right\|_1 \leq \max\left\{ \Lambda^{-1}(\delta^{-1} \max\{0, (\delta' - \delta)\rho_l^{\delta'}(\boldsymbol{X})\}), \Lambda^{-1}(\delta'^{-1} \max\{0, (\delta - \delta')\rho_l^{\delta}(\boldsymbol{X})\}) \right\}. \quad (6.2)$$

Part (i) of the theorem indicates that ρ_l^{δ} is locally Lipschitz continuous in δ and for fixed $\delta > 0$ a small perturbation of δ' from δ will not incur significant change of the relative change of the MLRM in that $|\rho_l^{\delta'}(\boldsymbol{X}) - \rho_l^{\delta}(\boldsymbol{X})|/|\rho_l^{\delta}(\boldsymbol{X})|$ is bounded by about $\frac{|\delta' - \delta|}{\delta}$. Moreover, it is important to note that while our quantitative result does not hold for every positive perturbation of $\delta = 0$, we can still obtain the similar result with a slight modification of the proof by leveraging Robinson's error bound theorem (Robinson, 1975). Part (ii) means that the optimal solution is stable w.r.t. small perturbation of δ. Note that inequalities (6.1) and (6.2) do not depend on Δ. The condition $\mathcal{S}^{\delta'}(\boldsymbol{X}) \cap B(\mathbf{0}, \Delta) \neq \emptyset$ simply means that problem (2.1) has a uniformly bounded optimal solution for all δ' close to δ.

6.2 Impact of exogenous random shock

We now turn to discuss the impact of a perturbation of the risky financial system $\boldsymbol{X}$ on the MLRM and the associated allocation.

Theorem 6.2 (Sensitivity of MLRM w.r.t. data perturbation) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function and $\delta > 0$ be a prespecified risk tolerance level. Assume: (a) l is Lipschitz continuous with modulus L, i.e., $|l(\boldsymbol{x}) - l(\boldsymbol{y})| \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_1$. (b) there exists a positive constant Δ such that $\mathcal{S}^{\delta}(\boldsymbol{Y}) \cap B(\mathbf{0}, \Delta) \neq \emptyset$ for all $\boldsymbol{Y}$ in a neighborhood of $\boldsymbol{X}$ equipped with Luxemburg norm $\|\cdot\|_{\theta}$. Then the following assertions hold.*

(i) For every $\boldsymbol{X}, \boldsymbol{Y} \in L_n^{\theta}$ and $\delta > 0$,

$$|\rho_l^{\delta}(\boldsymbol{X}) - \rho_l^{\delta}(\boldsymbol{Y})| \leq \frac{L}{\delta}\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1]. \quad (6.3)$$

(ii) Assume in addition: (c) Assumption 5.1 holds, (d) $l(\cdot)$ is strictly convex. Define the growth function

$$\Lambda(\eta) := \min \left\{ \sum_{i=1}^{n} \tau_i - \sum_{i=1}^{n} \tau_i^{\delta}(\boldsymbol{X}) : \mathsf{dl}_1(\boldsymbol{\tau}, \boldsymbol{\tau}^{\delta}(\boldsymbol{X})) \geq \eta, \ \forall \boldsymbol{\tau} \in B(\mathbf{0}, \Delta) \right\},$$

where $\boldsymbol{\tau}^{\delta}(\boldsymbol{X})$ is the set of optimal solutions to the problem (2.1) with $\boldsymbol{X}$, and $\Lambda^{-1}(\nu) := \sup\{\eta : \Lambda(\eta) \leq \nu\}$. Then

$$\|\boldsymbol{\tau}^{\delta}(\boldsymbol{X}) - \boldsymbol{\tau}^{\delta}(\boldsymbol{Y})\| \leq \Lambda^{-1}\left(\frac{L}{\delta}\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1]\right). \quad (6.4)$$

Inequality (6.3) implies that the MLRM is globally Lipschitz continuous over the Orlicz space L_n^θ, which resembles the Lipschitz continuity of law invariant coherent risk measure on L^p established by Inoue (2003, Lemma 2.1). It provides a limit for the rate of change of MLRM with respect to the change of X under $\|\cdot\|_1$-norm. Inequality (6.4) means that the impact of the perturbation of X on the optimal risk capital allocation is containable.

6.3 Resilence against data perturbation

In practice, obtaining the distribution of the random vector $\boldsymbol{X}$ can be a challenging task. Rather, it is often more feasible to obtain samples of the random vector, e.g., via historical data or Monte Carlo sampling. Let $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ denote the independent and identically distributed (i.i.d. for short) copies of $\boldsymbol{X}$. Then its empirical distribution is

$$P_N(\boldsymbol{x}) := \frac{1}{N}\sum_{j=1}^N \mathbb{1}_{\{\boldsymbol{x}^j \le \boldsymbol{x}\}},\ \forall \boldsymbol{x} \in \mathbb{R}^n.$$

Here and later on $\mathbb{1}_A$ denotes the indicator function of event A. Instead of solving problem (2.1), we can solve the associated sample average approximation (SAA) problem:

$$\text{(MLRM-SAA)} \quad \inf_{\boldsymbol{\tau} \in \mathbb{R}^n} \quad \sum_{i=1}^n \tau_i \tag{6.5a}$$

$$\text{s.t.} \quad \mathbb{E}_{P_N}[l(\boldsymbol{X} - \boldsymbol{\tau})] \le \delta \sum_{i=1}^n \tau_i. \tag{6.5b}$$

Let $\varrho_l^\delta(P_N)$ denote the optimal value of problem (6.5). Since the empirical distribution P_N is uniquely determined by the sample data $\{\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N\}$, then we can regard $\varrho_l^\delta(\cdot)$ as a mapping from $\mathbb{R}^{\otimes N}$ to $\mathbb{R}$. In other words, $\varrho_l^\delta(P_N) = \varrho_l^\delta(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N) =: \widehat{\varrho}_N$ is an estimator of $\rho_l^\delta(\boldsymbol{X})$.

In data-driven problems, however, sample data may be contaminated. In that case, we might regard the samples as being generated by perturbed probability distribution Q. Let $\{\widetilde{\boldsymbol{x}}^1, \ldots, \widetilde{\boldsymbol{x}}^N\}$ be i.i.d. copies $\boldsymbol{X}$ with distribution Q. Then the empirical distribution function is

$$Q_N(\boldsymbol{x}) := \frac{1}{N}\sum_{j=1}^N \mathbb{1}_{\{\widetilde{\boldsymbol{x}}^j \le \boldsymbol{x}\}},\ \forall \boldsymbol{x} \in \mathbb{R}^n, \tag{6.6}$$

and the statistical estimator of $\rho_l^\delta(\boldsymbol{X})$ is $\varrho_l^\delta(Q_N) = \varrho_l^\delta(\widetilde{\boldsymbol{x}}^1, \ldots, \widetilde{\boldsymbol{x}}^N) := \hat{\varrho}_N$, which is obtainable whereas $\varrho_l^\delta(P_N)$ is not. We are interested in whether $\varrho_l^\delta(Q_N)$ is a reasonably good estimator of $\rho_l^\delta(\boldsymbol{X})$ by comparing the laws of $\varrho_l^\delta(Q_N)$ and $\varrho_l^\delta(P_N)$ under some metrics.

To this end, we adopt the widely used Kantorovich/Wasserstein metric. Let $\mathcal{H}$ be the set of functions $h : \mathbb{R}^n \to \mathbb{R}$ satisfying $|h(\boldsymbol{x}) - h(\boldsymbol{y})| \le \|\boldsymbol{x} - \boldsymbol{y}\|_1$ and P, Q be two probability measures on $\mathbb{R}^n$. Recall that the Kantorovich metric between P and Q, denoted by $\mathsf{dl}_K(P, Q)$, is defined by

$$\mathsf{dl}_K(P, Q) := \sup_{h \in \mathcal{H}} \left\{ \int_{\mathbb{R}^n} h(\boldsymbol{x}) P(d\boldsymbol{x}) - \int_{\mathbb{R}^n} h(\boldsymbol{x}) Q(d\boldsymbol{x}) \right\}.$$

see, e.g., Gibbs and Su (2002). An important property of the Kantorovich metric is that it metrizes weak convergence of probability measures when the support set is bounded, that is, a sequence of probability measures $\{P_N\}$ converges to P weakly if and only if $\mathsf{dl}_K(P_N, P) \to 0$ as N tends to infinity, see, e.g., Billingsley (2013).

Theorem 6.3 (Quantitative statistical robustness) *Let $l : \mathbb{R}^n \mapsto \mathbb{R}$ be a normalized multivariate loss function satisfying Assumption 3.1 and $\delta > 0$ be a prespecified risk tolerance level. Define*

$$\mathcal{M}_n^l := \left\{ P \in \mathscr{P}(\mathbb{R}^n) : \int_{\mathbb{R}^n} l(|\boldsymbol{t}|) P(d\boldsymbol{t}) < \infty \right\}.$$

If l is Lipschitz continuous with modules L, then for any $N \geq 1$

$$\mathsf{dl}_K \left(P^{\otimes N} \circ \widehat{\varrho}_N^{-1}, Q^{\otimes N} \circ \widehat{\varrho}_N^{-1} \right) \leq \frac{L}{\delta} \mathsf{dl}_K(P, Q),\ \forall P, Q \in \mathcal{M}_n^l, \tag{6.7}$$

where $P^{\otimes N} \circ \widehat{\varrho}_N^{-1}$ is a probability measure over $\mathbb{R}$ induced by random variable $\widehat{\varrho}_N(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N)$ where $(\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N)$ follows joint probability distribution $P^{\otimes N} := \underbrace{P \times \cdots \times P}_{N \text{ times}}$ over $(\mathbb{R}^n)^{\otimes N} := \underbrace{\mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{N \text{ times}}$.

Theorem 6.3 means that the difference between the probability distribution of $\varrho_l^\delta(P_N)$ and the probability distribution of $\varrho_l^\delta(Q_N)$ under the Kantorovich metric is linearly bounded by $\mathsf{dl}_K(Q, P)$. The theorem ensures that the performance of $\varrho_l^\delta(P_N)$ is close to that of $\varrho_l^\delta(Q_N)$ when Q is close to P. In other words, if the uncertainty data are within the specific structure $(P, Q \in \mathcal{M}_n^l)$, then we are guaranteed that $\varrho_l^\delta(Q_N)$ is a good estimate of $\varrho_l^\delta(P_N)$. This kind of guarantee is also useful when we regard Q_N as the empirical distribution based on future uncertainty data whereas P_N is the empirical distribution based on present sample data. In that case, $\varrho_l^\delta(P_N)$ is the MLRM for future and the theorem ensures the quantity is close to $\varrho_l^\delta(Q_N)$ obtained with the present sample data.

7 Empirical studies

In this section, we examine the performance of the proposed MLRM and associated optimal risk capital allocation. All of the experiments are carried out on MATLAB 2018b installed on a Macbook Pro (i5-5257 CPU, 2.90GHz dual core processor, 8GB memory).

7.1 Setup

Consider a risky financial system with $n = 10$ risky components. For simplicity, we assume that $\boldsymbol{X} = (X_1, \ldots, X_{10})^{\mathrm{T}}$ is a normal random vector with normal margins $X_i \sim \mathcal{N}(i \times 3\%, i \times 2.5\%)$ for $i = 1, \ldots, n$ and a specified common pairwise correlation $\rho \in \{-0.1, 0, 0.5\}$. For each case of dependence, we randomly generate $N = 1000$ realizations of $\boldsymbol{X}$ as the required panel data. Moreover, to analyze the impact of different multivariate loss function on the proposed measure and associated allocation, we consider the following normalized multivariate loss functions from Example 2.1:

$$l(\boldsymbol{x}) := \alpha\left[\left(\sum_{i=1}^{n} x_i\right)_+ - \beta_0\left(\sum_{i=1}^{n} x_i\right)_-\right] + (1-\alpha)\left[\sum_{i=1}^{n}(x_i)_+ - \sum_{i=1}^{n}\beta_i(x_i)_-\right], \tag{7.1}$$

where $\alpha \in [0,1]$ and $0 \leq \beta_i < 1$ for $i = 0, 1, \ldots, n$. The multivariate loss function (7.1) considers both the aggregate risk and individual risk as well as the risk preferences of a decision maker. Specifically, the term $(\sum_{i=1}^{n} x_i)_+$ (or $(\sum_{i=1}^{n} x_i)_-$) captures the loss (or the gain) of the aggregate risk and $\sum_{i=1}^{n}(x_i)_+$ (or $\sum_{i=1}^{n}(x_i)_-$) represents the aggregate loss (or the aggregate gain) of all individual risks. In addition, the parameter α represents the relative importance of the corresponding risk or the decision maker's preferences on the corresponding risk. Note that $f = f_+ - f_-$, then (7.1) can be rewritten as

$$l(\boldsymbol{x}) = \alpha(1-\beta_0)\left(\sum_{i=1}^{n} x_i\right)_+ + \alpha\beta_0\sum_{i=1}^{n} x_i + (1-\alpha)\sum_{i=1}^{n}(1-\beta_i)(x_i)_+ + (1-\alpha)\sum_{i=1}^{n}\beta_i x_i. \tag{7.2}$$

Consequently, the multivariate loss ratio can be reformulated as

$$\inf_{\boldsymbol{\tau};\boldsymbol{\eta},\boldsymbol{\nu}} \quad \sum_{i=1}^{n}\tau_i \tag{7.3a}$$

$$\text{s.t.} \quad \frac{\alpha(1-\beta_0)}{N}\sum_{j=1}^{N}\eta_j + \frac{1-\alpha}{N}\sum_{j=1}^{N}\sum_{i=1}^{n}(1-\beta_i)\nu_{ij} - \sum_{i=1}^{n}(\alpha\beta_0 + (1-\alpha)\beta_i + \delta)\tau_i$$
$$\leq -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{n}(\alpha\beta_0 + (1-\alpha)\beta_i)X_{ij}, \tag{7.3b}$$

$$-\eta_j - \sum_{i=1}^{n}\tau_i \leq -\sum_{i=1}^{n}X_{ij},\ j = 1, \ldots, N, \tag{7.3c}$$

$$-\nu_{ij} - \tau_i \leq -X_{ij},\ i = 1, \ldots, n; j = 1, \ldots, N, \tag{7.3d}$$

$$\eta_j \geq 0; \nu_{ij} \geq 0, \tag{7.3e}$$

where $X_{ij} = X_i(\omega_j)$ is the j-th realization of X_i. Note that (7.3) is a linear program and can be easily solved by using *linprog* in Matlab.

7.2 Experiments

We consider the following experiments to investigate the performance of the proposed measure:

Experiment 1: Comparative statistics for multivariate loss functions.

In this experiment, we investigate the impact of the parameters in multivariate loss function (7.1) on the value of proposed measure and its associated allocation. Specifically, we solve the linear program (7.3) using the dataset randomly generated for the common pairwise correlation $\rho = 0.5$, the risk tolerance level $\delta = 0.01$, and the specified parameters α, β_0 and β_i for $i = 1, \ldots, n$. Figure 2 depicts the value of multivariate loss ratio measure $\rho_l^\delta(\boldsymbol{X})$ and its corresponding optimal allocation $\boldsymbol{\tau}^*$ as a function of the parameters in multivariate loss function (7.1). From Figure 2 (a) and (b), we observe that, for fixed δ and $\beta_0 = \beta_i$, $\rho_l^\delta(\boldsymbol{X})$ exhibits a decreasing tread as α increases from 0 to 0.9. This is because a higher

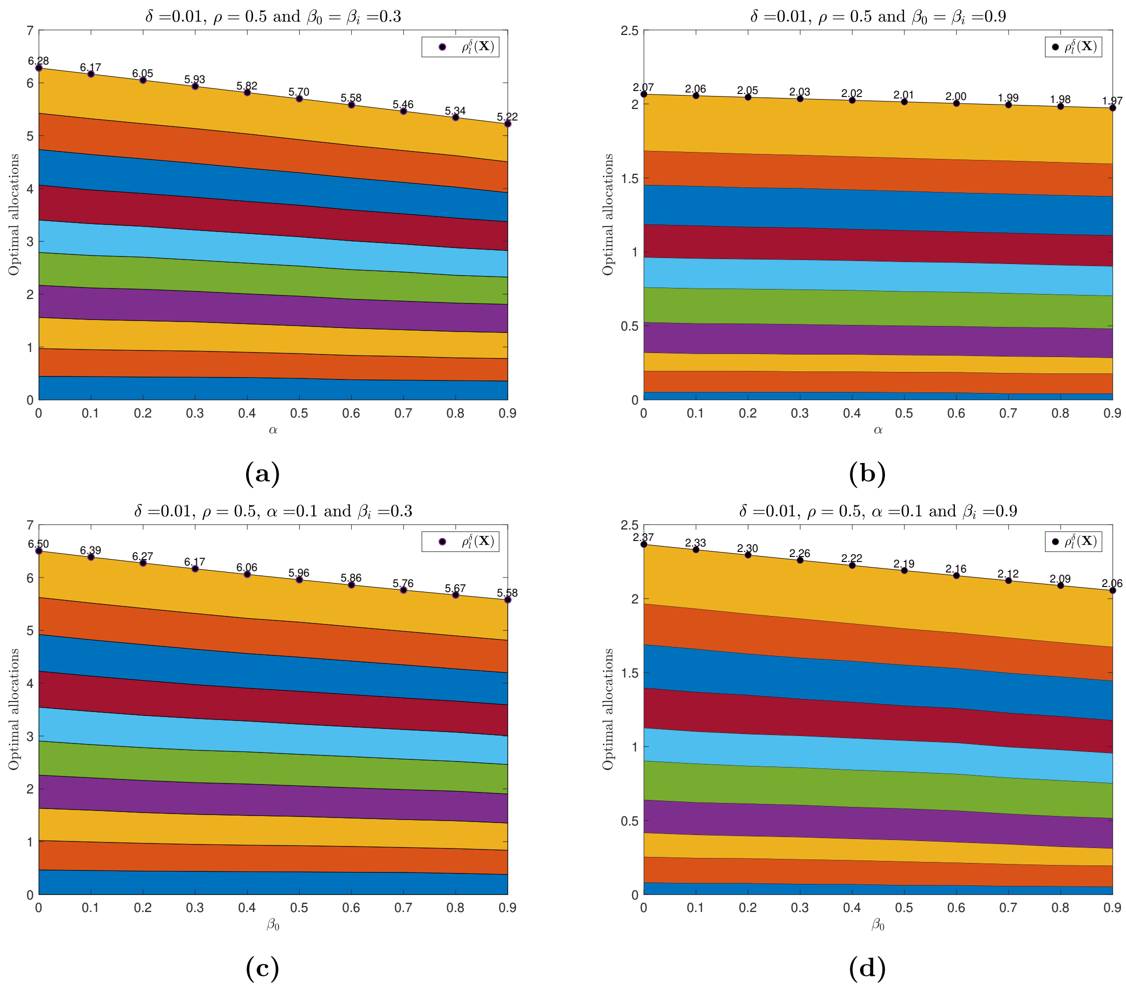


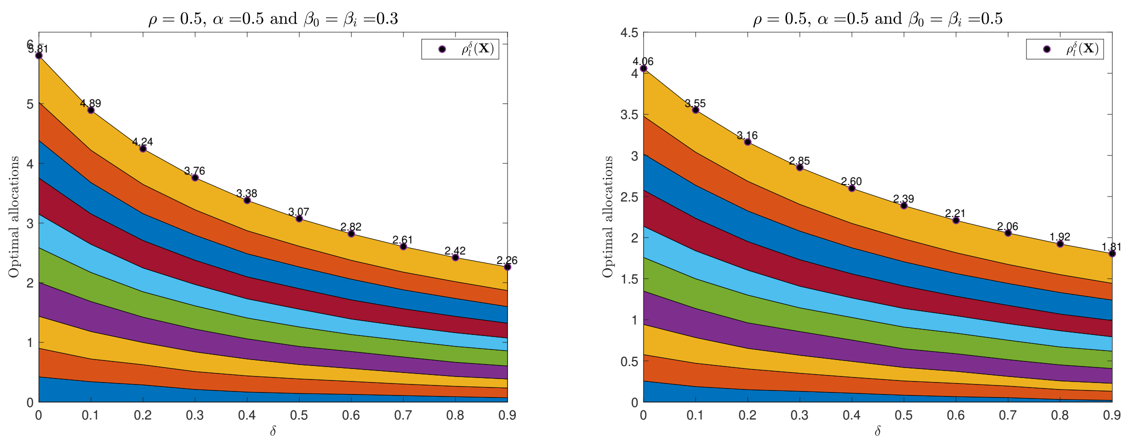
Figure 2: The value of MLRM and its optimal allocation composition as a function of the parameters in multivariate loss function l with $\delta = 0.01$ and $\rho = 0.5$. **(a)** $\beta_0 = \beta_i = 0.3$, **(b)** $\beta_0 = \beta_i = 0.9$, **(c)** $\alpha = 0.1, \beta_i = 0.3$, and **(d)** $\alpha = 0.1, \beta_i = 0.9$. The optimal allocations $(\tau_1^*, \ldots, \tau_{10}^*)$ are depicted in ascending order with coloured stripes, i.e., the allocation to X_{10} at the top and that of X_9 second from top, and so on. The dark dotted points indicate the values of MLRM at different δ values ($\rho_l^\delta(\boldsymbol{X}) = \sum_{i=1}^n \tau_i^*$).

α corresponds to a decision maker assigning greater significance to the net loss of aggregate risk, as indicated by the first term in (7.1). Consequently, $l(\boldsymbol{x})$ decreases as α increases, given that the loss of aggregate risk is less than the aggregate loss of individual risks, as shown in equation (7.2). Thus, $\rho_l^\delta(X)$ decreases as α increases since the feasible set to problem (2.1) becomes larger. Furthermore, it is evident that the largest allocation of risk capital $\rho_l^\delta(\boldsymbol{X})$ is made to X_{10} due to its highest mean and variance.

From Figure 2 (c) and (d), we see that $\rho_l^\delta(\boldsymbol{X})$ decreases as β_0 increases. This is because a larger β_0 corresponds to a higher proportionate subsidy of the loss of aggregate risk, and consequently, less risk capital is needed to ensure that the relative multivariate shortfall risk is below the specified tolerance level δ.

Experiment 2: Comparative statistics for risk tolerance level.

In this experiment, we investigate the impact of the risk tolerance level δ on MLRM value $\rho_l^\delta(\boldsymbol{X})$ and the corresponding allocation of the risk capital. To this end, we consider a fixed multivariate loss function l as defined in (7.1) with $\alpha = 0.5$ and $\beta_0 = \beta_i = 0.3$ for $i = 1, \ldots, n$ and with $\alpha = 0.5$ and $\beta_0 = \beta_i = 0.5$ for $i = 1, \ldots, n$. Figure 3 depicts the values of MLRM $\rho_l^\delta(\boldsymbol{X})$ and the corresponding optimal allocations



(a) **(b)**

Figure 3: The value of MLRM and the related optimal allocation composition as a function of the risk tolerance level δ with $\alpha = 0.5$ and $\rho = 0.5$. **(a)** $\beta_0 = \beta_i = 0.3$ and **(b)** $\beta_0 = \beta_i = 0.5$. The optimal allocations are depicted in ascending order with coloured stripes, i.e., the allocation to X_{10} at the top and that of X_9 second from top, and so on. The dark dotted points indicate the values of MLRM at different δ values. In the case that $\delta = 0$, MLRM coincides with the MSRM by Armenti et al. (2018).

$\boldsymbol{\tau}^*$ as a function of the risk tolerance level δ. In Figure 3 (a) and (b), we can observe that, in both cases, $\rho_l^\delta(\boldsymbol{X})$ decreases when δ increases from 0 to 0.9 in that less risk capital is required to ensure that the relative shortfall risk remains below a higher risk tolerance level δ. Moreover, in both cases, as δ increases, the allocation weight assigned to X_{10} also increases, while the allocation weight assigned to X_1 decreases. This shift occurs because, with a higher risk tolerance level δ, the decision maker places greater importance on mitigating the largest risk as opposed to the smallest risk. The results also show that MLRM and the related allocation are stable w.r.t. small change of δ as envisaged by Theorem 6.1. To demonstrate the quantitative results of Theorem 6.1, we consider the case that $\delta' \geq \delta > 0$. From

Theorem 6.1 and the positiveness of $\rho_l^\delta(\boldsymbol{X})$, we have

$$\frac{\rho_l^\delta(\boldsymbol{X}) - \rho_l^{\delta'}(\boldsymbol{X})}{\rho_l^{\delta'}(\boldsymbol{X})} \leq \frac{\delta' - \delta}{\delta} \iff \delta' \rho_l^{\delta'}(\boldsymbol{X}) - \delta \rho_l^\delta(\boldsymbol{X}) \geq 0 \text{ if } \delta' \geq \delta > 0.$$

Consequently, the mapping $\delta \mapsto \delta \rho_l^\delta(\boldsymbol{X})$ is non-decreasing for any $\delta > 0$. Thus, in Experiment 2, it might be better to test the non-decreasing property of $\delta \rho_l^\delta(\boldsymbol{X})$ for $\delta > 0$. Figure 4 depicts $\delta \rho_l^\delta(\boldsymbol{X})$ as a function of the risk tolerance level δ with $\rho = 0.5$ and $\alpha = 0.5$. As shown in the figure, the function $\delta \rho_l^\delta(\boldsymbol{X})$ is indeed non-decreasing in δ, which confirms the theoretical result in Theorem 6.1.

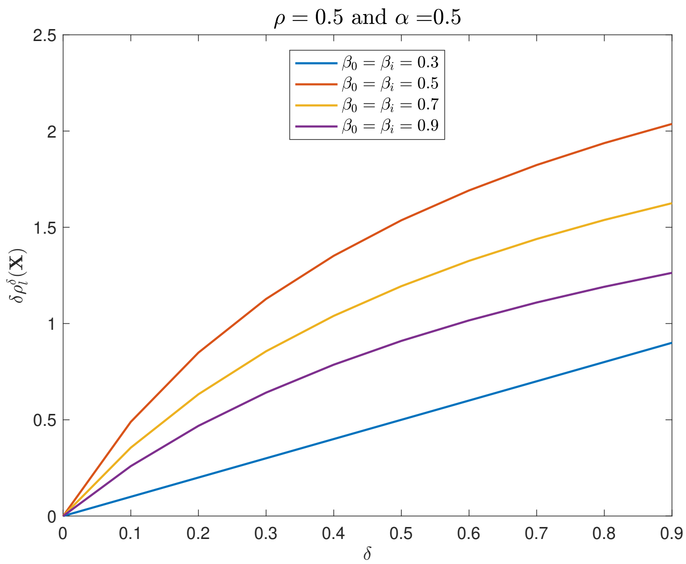
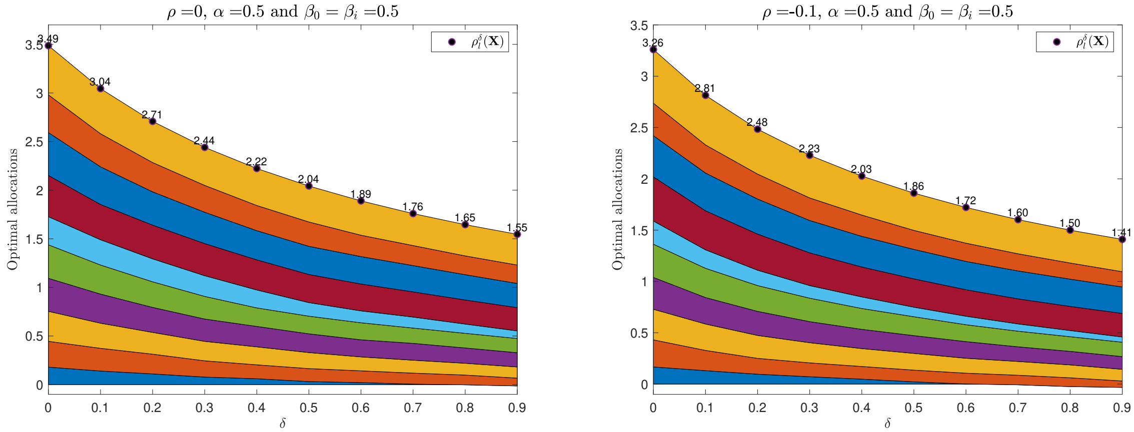


Figure 4: The value of $\delta \rho_l^\delta(\boldsymbol{X})$ as a function of the risk tolerance level δ with $\rho = 0.5$ and $\alpha = 0.5$.

Experiment 3: Comparative statistics for data dependence.

In Example 5.3, we discuss the effect of data dependence on the value of MLRM and related allocation of risk capital when the loss function has a relatively simple structure. In this experiment, we take a step further to give some numerical analysis about the impact of data dependence on the quantities when the loss function takes a slightly more complex form as defined in (7.1). As in the previous experiment, we set $\alpha = 0.5$ and $\beta_0 = \beta_i = 0.5$ and allow the risk tolerance level δ to vary from 0 to 0.9. Figure 5 depicts the values of MLRM $\rho_l^\delta(\boldsymbol{X})$ and its corresponding optimal allocation $\boldsymbol{\tau}^*$ as a function of the risk tolerance level δ. From Figure 5(a)-(b) and Figure 3(b), it is evident that, for any $\delta \in [0, 0.9]$, $\rho_l^\delta(\boldsymbol{X})$ is largest when the data exhibits common pairwise positive correlation, while $\rho_l^\delta(\boldsymbol{X})$ is smallest in the case of negative correlation. This difference is primarily due to the presence of diversification benefits in the negative correlation case. As expected, the figure demonstrates that the proposed measure is sensitive to the levels of interdependence between components. Moreover, in both cases, as δ increases, the allocation weight assigned to X_{10} also increases, while the allocation weight assigned to X_1 decreases. This shift occurs because, with a higher risk tolerance level δ, the decision maker places greater importance on mitigating the largest risk as opposed to the smallest risk.

Experiment 4: Robustness check to data perturbation.



(a) **(b)**

Figure 5: The value of MLRM and the optimal allocation composition as a function of the risk tolerance level δ with $\alpha = 0.5$ and $\beta_0 = \beta_i = 0.5$. **(a)** $\rho = 0$ and **(b)** $\rho = -0.1$. The optimal allocations are depicted in ascending order, i.e., the allocation to X_{10} at the top and that of X_9 second from top, and so on. Note that the allocation to X_1 becomes negative when δ exceeds 0.8 in (a) and 0.7 in (b).

In this experiment, we study stability of the MLRM and related optimal allocation. Specifically, we perturb the original dataset randomly up to $\pm 50\%$ in terms of their values. For example, the j-th realization X_{ij} of X_i in the contaminated dataset, denoted by $\widetilde{X}_{ij}$, is given by

$$\widetilde{X}_{ij} = X_{ij} \times (1 + r \times s),$$

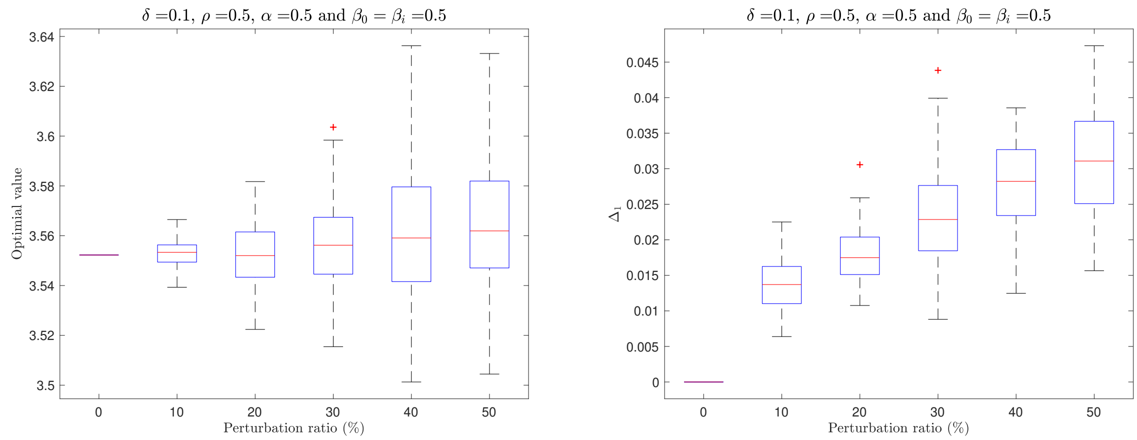
where r is a perturbation ratio (we choose $r = 5\%$ in our experiments) and s is drawn from the uniform distribution $[-1, 1]$. In this experiment, we fix the multivariate loss function l with $\alpha = 0.5$ and $\beta_0 = \beta_i = 0.5$ for $i = 1, \ldots, n$. We set the pairwise correlation parameter $\rho = 0.5$ when generating the original dataset and specify the risk tolerance level $\delta = 0.1$. Let $\boldsymbol{\tau}^*$ and $\widetilde{\boldsymbol{\tau}}^*$ denote the optimal allocation based on the original dataset and perturbed dataset, respectively. Define

$$\Delta_1 = \left\| \frac{\widetilde{\boldsymbol{\tau}}^*}{\rho_l^\delta(\widetilde{\boldsymbol{X}})} - \frac{\boldsymbol{\tau}^*}{\rho_l^\delta(\boldsymbol{X})} \right\|_1$$

as the difference of the normalized optimal allocations based on the original dataset and perturbed dataset. We randomly generate 100 perturbed dataset for each $r \in \{0, 10\%, 20\%, 30\%, 40\%, 50\%\}$ and solve the model (7.3) to obtain the optimal value and optimal solutions for each perturbed dataset. Figure 6 (a) and (b) depicts boxplot of MLRM values and the difference (Δ_1) of normalized allocations as r increases, respectively. From the figure, we can see that the change of the two quantities are not drastic, which shows robustness of our model to data perturbation as envisaged in Theorem 6.2.

Experiment 5: Statistical robustness of MLRM.

In this experiment, we study the statistical robustness of MLRM. To ease the computation, we regard the empirical distribution of the original data set as the true distribution, denoted by P, and that of the perturbed data set generated in Experiment 4 as the perturbed distribution, denoted by Q. Next, we use the two discrete distributions (P and Q) to randomly generate samples respectively with fixed sample size $N = 100$ and then compute respective MLRMs by solving problem (7.3) based on the generated



(a) **(b)**

Figure 6: **(a)** Boxplot of the optimal value of MLRM, and **(b)** Boxplot of the distance of optimal allocation weights (Δ_1), w.r.t. the perturbation ratio (%) with $\delta = 0.1$, $\alpha = 0.5$ and $\beta = \beta_i = 0.5$.

empirical distribution P_N and Q_N (with $N = 100$). We construct the respective approximated CDFs after running 1000 simulations. Figure 7 depicts the CDFs of the simulated results. From the figure, we

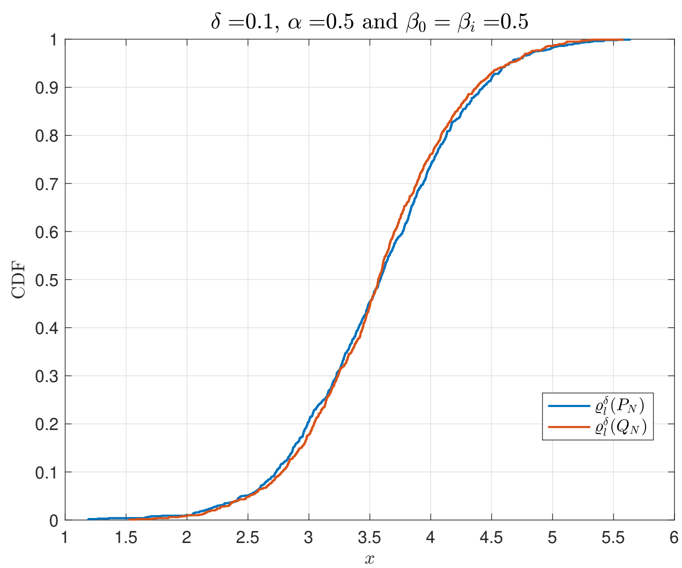


Figure 7: The CDFs of $\varrho_l^\delta(P_N)$ and $\varrho_l^\delta(Q_N)$.

can see that the discrepancy between the CDFs of $\varrho_l^\delta(P_N)$ and $\varrho_l^\delta(Q_N)$ is small which means that MLRM is statistically robust. To verify the error bound provided in Theorem 6.3, we calculate the difference between P and Q and the difference between $\varrho_l^\delta(P_N)$ and $\varrho_l^\delta(Q_N)$ under the Kantorovich metric:

$$\mathsf{dl}_K(P, Q) = 0.1972 \quad \text{and} \quad \mathsf{dl}_K(\varrho_l^\delta(P_N), \varrho_l^\delta(Q_N)) = 0.0423,$$

which satisfies (6.7) because in this experiment, we set $\delta = 0.1$ and the multivariate loss function (7.1) is Lipschitz continuous with modulus $L = 1$ ($L/\delta = 10 > 0.0423/0.1972$).

8 Concluding remarks

In this paper, we propose a new notion, called MLRM, which may be viewed as an extension of the loss ratio of Baron et al. (2023) to a multivariate setting and a generalization of the MSRM of Armenti et al. (2018) to a broader framework which allows one to balance the expected systemic risk to be reduced and the amount of the capital to be injected. Under some moderate conditions, we show that the MLRM enjoys a number of desirable properties of the MSRM and the loss ratio. As a main motivation and application, we apply the MLRM to risk capital allocation problems and examine the similarities and differences between MLRM and MSRM in terms of capital commitment and efficiency. Moreover, we carry out comprehensive sensitivity analysis with regard to variation of the degree of risk tolerance and uncertainty data perturbation. These analysis provides theoretical guidance for practical application of the MLRM particularly in data-driven problems.

This work may be extended in several directions. First, our analysis of the main properties of the MLRM is focused on the case that the risk tolerance parameter (δ) is restricted to be non-negative. It might be interesting to discuss the case that δ takes negative values because the MLRM would recover an important optimal risk capital allocation model with cost of the risk capital (Laeven and Goovaerts, 2004; Wang et al., 2023). It will also be interesting to extend the model which weighs investments differently, or under alternative capital aggregation functions which are not necessarily additive, see e.g. Noyan and Rudolf (2013) for multivariate CVaR. Second, since the paper is mainly focused on risk capital allocation, we have not discussed how MLRM may be used in multi-resources allocation problems although this will be very interesting. In particular, when the loss function is ambiguous, one may consider a preference robust model as in Delage et al. (2022). Third, the proposed MLRM may be applied to multistage decision making problems where the balance between shortfall risk and capital input is adjusted at different stages. We leave all these for future research.

Acknowledgments.

We would like to thank three anonymous referees for valuable comments which help us strengthen the paper. We also thank the Associate Editor for effective handling of the review. Wei Wang acknowledges financial support from the National Natural Science Foundation of China (No. 12501456) and the Fundamental Research Funds for the Central Universities.

References

Acharya, V. V., Pedersen, L. H., Philippon, T., and Richardson, M. (2017). Measuring systemic risk. *The review of financial studies*, 30(1):2–47.

Adrian, T. and Brunnermeier, M. K. (2016). Covar. *The American Economic Review*, 106(7):1705.

Ararat, Ç., Hamel, A. H., and Rudloff, B. (2017). Set-valued shortfall and divergence risk measures. *International Journal of Theoretical and Applied Finance*, 20(05):1750026.

Armbruster, B. and Delage, E. (2015). Decision making under uncertainty when preference information is incomplete. *Management science*, 61(1):111–128.

Armenti, Y., Crépey, S., Drapeau, S., and Papapantoleon, A. (2018). Multivariate shortfall risk allocation and systemic risk. *SIAM Journal on Financial Mathematics*, 9(1):90–126.

Balog, D., Bátyi, T. L., Csóka, P., and Pintér, M. (2017). Properties and comparison of risk capital allocation methods. *European Journal of Operational Research*, 259(2):614–625.

Baron, O., Berman, O., Cire, A. A., and Roshanaei, V. (2023). Loss ratio as a risk constraint: Analysis and optimization. *Available at SSRN 4425015*.

Bawa, V. S. and Lindenberg, E. B. (1977). Capital market equilibrium in a mean-lower partial moment framework. *Journal of financial economics*, 5(2):189–200.

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.

Biagini, F., Fouque, J.-P., Frittelli, M., and Meyer-Brandis, T. (2019). A unified approach to systemic risk measures via acceptance sets. *Mathematical Finance*, 29(1):329–367.

Biagini, F., Fouque, J.-P., Frittelli, M., and Meyer-Brandis, T. (2020). On fairness of systemic risk measures. *Finance and Stochastics*, 24(2):513–564.

Billingsley, P. (2013). *Convergence of probability measures*. John Wiley & Sons.

Bisias, D., Flood, M., Lo, A. W., and Valavanis, S. (2012). A survey of systemic risk analytics. *Annu. Rev. Financ. Econ.*, 4(1):255–296.

Bonnans, J. F. and Shapiro, A. (2013). *Perturbation analysis of optimization problems*. Springer Science & Business Media.

Brown, D. B., Giorgi, E. D., and Sim, M. (2012). Aspirational preferences and their representation by risk measures. *Management Science*, 58(11):2095–2113.

Brown, D. B. and Sim, M. (2009). Satisficing measures for analysis of risky positions. *Management Science*, 55(1):71–84.

Brownlees, C. and Engle, R. F. (2017). Srisk: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies*, 30(1):48–79.

Brunnermeier, M. K. and Cheridito, P. (2019). Measuring and allocating systemic risk. *Risks*, 7(2):46.

Cai, J., Jia, H., and Mao, T. (2022). A multivariate cvar risk measure from the perspective of portfolio risk management. *Scandinavian Actuarial Journal*, 2022(3):189–215.

Castagnoli, E., Cattelan, G., Maccheroni, F., Tebaldi, C., and Wang, R. (2022). Star-shaped risk measures. *Operations Research*, 70(5):2637–2654.

Chen, C., Iyengar, G., and Moallemi, C. C. (2013). An axiomatic approach to systemic risk. *Management Science*, 59(6):1373–1388.

Cheridito, P. and Li, T. (2009). Risk measures on orlicz hearts. *Mathematical Finance: An International Journal of Mathematics, Statistics and Financial Economics*, 19(2):189–214.

Chu, Y., Deng, S., and Xia, C. (2020). Bank geographic diversification and systemic risk. *The Review of Financial Studies*, 33(10):4811–4838.

Clemen, R. T. and Reilly, T. (2013). *Making hard decisions with DecisionTools*. Cengage Learning.

Delage, E., Guo, S., and Xu, H. (2022). Shortfall risk models when information on loss function is incomplete. *Operations Research*, 70(6):3511–3518.

Doldi, A., Feng, Y., Fouque, J.-P., and Frittelli, M. (2023). Multivariate systemic risk measures and computation by deep learning algorithms. *Quantitative Finance*, 0(0):1–14.

Doldi, A. and Frittelli, M. (2021). Conditional systemic risk measures. *SIAM Journal on Financial Mathematics*, 12(4):1459–1507.

Doldi, A. and Frittelli, M. (2022). Multivariate systemic optimal risk transfer equilibrium. *Annals of Operations Research*, pages 1–46.

Drapeau, S. and Kupper, M. (2013). Risk preferences and their robust representation. *Mathematics of Operations Research*, 38(1):28–62.

Eisenberg, L. and Noe, T. H. (2001). Systemic risk in financial systems. *Management Science*, 47(2):236–249.

Ellis, S., Sharma, S., and Brzeszczyński, J. (2022). Systemic risk measures and regulatory challenges. *Journal of Financial Stability*, 61:100960.

Feinstein, Z., Rudloff, B., and Weber, S. (2017). Measures of systemic risk. *SIAM Journal on Financial Mathematics*, 8(1):672–708.

Fishburn, P. C. (1977). Mean-risk analysis with risk associated with below-target returns. *The American Economic Review*, 67(2):116–126.

Föllmer, H. and Leukert, P. (1999). Quantile hedging. *Finance and Stochastics*, 3:251–273.

Föllmer, H. and Leukert, P. (2000). Efficient hedging: cost versus shortfall risk. *Finance and Stochastics*, 4:117–146.

Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. *Finance and stochastics*, 6:429–447.

Föllmer, H. and Schied, A. (2016). Stochastic finance. In *Stochastic Finance*. de Gruyter.

Gibbs, A. L. and Su, F. E. (2002). On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435.

Girardi, G. and Ergün, A. T. (2013). Systemic risk measurement: Multivariate garch estimation of covar. *Journal of Banking & Finance*, 37(8):3169–3180.

González-Ortega, J., Radovic, V., and Ríos Insua, D. (2018). Utility elicitation. *Elicitation: The science and art of structuring judgement*, pages 241–264.

Guo, Q., Bauer, D., and Zanjani, G. (2021). Capital allocation techniques: Review and comparison. *Variance*, 14.

Guo, S. and Xu, H. (2021). Statistical robustness in utility preference robust optimization models. *Mathematical Programming*, 190(1-2):679–720.

Haezendonck, J. and Goovaerts, M. (1982). A new premium calculation principle based on orlicz norms. *Insurance: Mathematics and Economics*, 1(1):41–53.

Harlow, W. V. and Rao, R. K. (1989). Asset pricing in a generalized mean-lower partial moment framework: Theory and evidence. *Journal of financial and quantitative analysis*, 24(3):285–311.

Hoffmann, H., Meyer-Brandis, T., and Svindland, G. (2016). Risk-consistent conditional systemic risk measures. *Stochastic Processes and their Applications*, 126(7):2014–2037.

Hoffmann, H., Meyer-Brandis, T., and Svindland, G. (2018). Strongly consistent multivariate conditional risk measures. *Mathematics and Financial Economics*, 12:413–444.

Inoue, A. (2003). On the worst conditional expectation. *Journal of Mathematical Analysis and Applications*, 286(1):237–247.

Keeney, R. L. and Raiffa, H. (1993). *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press.

Kromer, E., Overbeck, L., and Zilch, K. (2016). Systemic risk measures on general measurable spaces. *Mathematical Methods of Operations Research*, 84:323–357.

Laeven, R. J. and Goovaerts, M. J. (2004). An optimization approach to the dynamic allocation of economic capital. *Insurance: Mathematics and Economics*, 35(2):299–319.

Long, D. Z., Sim, M., and Zhou, M. (2023). Robust satisficing. *Operations Research*, 71(1):61–82.

López-Espinosa, G., Moreno, A., Rubia, A., and Valderrama, L. (2015). Systemic risk and asymmetric responses in the financial industry. *Journal of Banking & Finance*, 58:471–485.

Molchanov, I. and Cascos, I. (2016). Multivariate risk measures: a constructive approach based on selections. *Mathematical Finance*, 26(4):867–900.

Noyan, N. and Rudolf, G. (2013). Optimization with multivariate conditional value-at-risk constraints. *Operations research*, 61(4):990–1013.

Poledna, S., Martínez-Jaramillo, S., Caccioli, F., and Thurner, S. (2021). Quantification of systemic risk from overlapping portfolios in the financial system. *Journal of Financial Stability*, 52:100808.

Pomerol, J. C. (1981). The boundedness of the lagrange multipliers set and duality in mathematical programming. *Zeitschrift für Operations Research*, 25:191–204.

Richard, S. F. (1975). Multivariate risk aversion, utility independence and separable utility functions. *Management Science*, 22(1):12–21.

Rizwan, M. S., Ahmad, G., and Ashraf, D. (2020). Systemic risk: The impact of covid-19. *Finance Research Letters*, 36:101682.

Robinson, S. M. (1975). An application of error bounds for convex programming in a linear space. *SIAM Journal on Control*, 13(2):271–273.

Rockafellar, R. T. (1997). *Convex analysis*, volume 11. Princeton university press.

Roy, A. D. (1952). Safety first and the holding of assets. *Econometrica: Journal of the econometric society*, pages 431–449.

Rüschendorf, L. (2004). Comparison of multivariate risks and positive dependence. *Journal of Applied Probability*, 41(2):391–406.

Russell, W. R. and Seo, T. K. (1978). Ordering uncertain prospects: the multivariate utility functions case. *The Review of Economic Studies*, 45(3):605–610.

Ruszczyński, A. and Shapiro, A. (2006). Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452.

Sahin, C., de Haan, J., and Neretina, E. (2020). Banking stress test effects on returns and risks. *Journal of Banking & Finance*, 117:105843.

Scarsini, M. (1988). Dominance conditions for multivariate utility functions. *Management Science*, 34(4):454–460.

Veraart, L. A. M. (2020). Distress and default contagion in financial networks. *Mathematical Finance*, 30(3):705–737.

Wang, W., Xu, H., and Ma, T. (2021). Quantitative statistical robustness for tail-dependent law invariant risk measures. *Quantitative Finance*, 21(10):1669–1685.

Wang, W., Xu, H., and Ma, T. (2023). Optimal scenario-dependent multivariate shortfall risk measure and its application in risk capital allocation. *European Journal of Operational Research*, 306(1):322–347.

Weber, S. and Weske, K. (2017). The joint impact of bankruptcy costs, fire sales and cross-holdings on systemic risk in financial networks. *Probability, Uncertainty and Quantitative Risk*, 2:1–38.

Wu, Q., Zhang, S., Wang, W., and Xu, H. (2023). Multi-attribute utility preference robust optimization: A continuous piecewise linear approximation approach. *arXiv preprint arXiv:2303.16420*.

Zhang, Y., Xu, H., and Wang, W. (2022). Preference robust models in multivariate utility-based shortfall risk minimization. *Optimization Methods and Software*, 37(2):712–752.

Appendix A Some additional material

A.1 Multivariate loss functions

Example A.1 (Chen et al. (2013)) *Consider the multivariate aggregation function defined as*

$$\begin{aligned} l(\boldsymbol{x}) := \inf_{\boldsymbol{y},\boldsymbol{b}\in\mathbb{R}_+^n} \quad & \sum_{i=1}^n y_i + \gamma \sum_{i=1}^n b_i \\ \text{s.t.} \quad & b_i + y_i \geq x_i + \sum_{j=1}^n \Pi_{ji} y_j,\ i = 1, \ldots, n, \end{aligned}$$

where $\gamma > 1$ and $0 \leq \Pi_{ij} \leq 1$ satisfying $\sum_{j=1}^n \Pi_{ij} = 1$ for $i = 1, \ldots, n$. The interpretation of such a multivariate loss function is as follows: The parameter Π_{ij} denotes the fraction of the total debt of firm i that is owed to firm j. In the event of a loss x_i incurred by firm i, this loss must be covered either by firm i reducing the payments on its obligations to other firms by an amount y_i, or by seeking an injection of external funds from regulators in the amount b_i. The parameter γ plays a crucial role in balancing the trade-off between two essential factors: the aggregate shortfalls $\sum_{i=1}^n y_i$ arising from interfirm obligations within the economy and the cost $\sum_{i=1}^n b_i$ of injecting new capital to support the economy. By considering this tradeoff, the multivariate aggregation function effectively measures the net systemic cost of contagion.

Example A.2 (Doldi et al. (2023)) *Consider the case that the multivariate loss functions are generated by the multiattribute utility function,*

$$\begin{aligned} l(\boldsymbol{x}) = -u(-\boldsymbol{x}) &= \frac{1}{2}\sum_{i=1}^n (\exp(2\alpha_i x_i) - 1) + \frac{1}{2}\sum_{i,j=1;i\neq j}^n (\exp(\alpha_i x_i + \alpha_j x_j) - 1) \\ &= \frac{1}{2}\left(\sum_{i=1}^n \exp(\alpha_i x_i)\right)^2 - \frac{1}{2}n^2, \end{aligned}$$

where $\alpha_i > 0$ for $i = 1, \ldots, n$. The interpretation of such a multivariate loss function is that it takes into account not only the individual's loss but also the aggregate pairwise losses among the different components. It could be used if a decision maker prefers multiple small losses over one large loss.

A.2 Some illustration examples of MLRM

Example A.3 (Illustrations of MLRM) *Consider loss function $l_1(x_1, x_2) = 2(x_1 - 1)_+ + 2x_2$. In this case, we can deduce from the feasible set depicted in Figure A.1 (a) and (b) that*

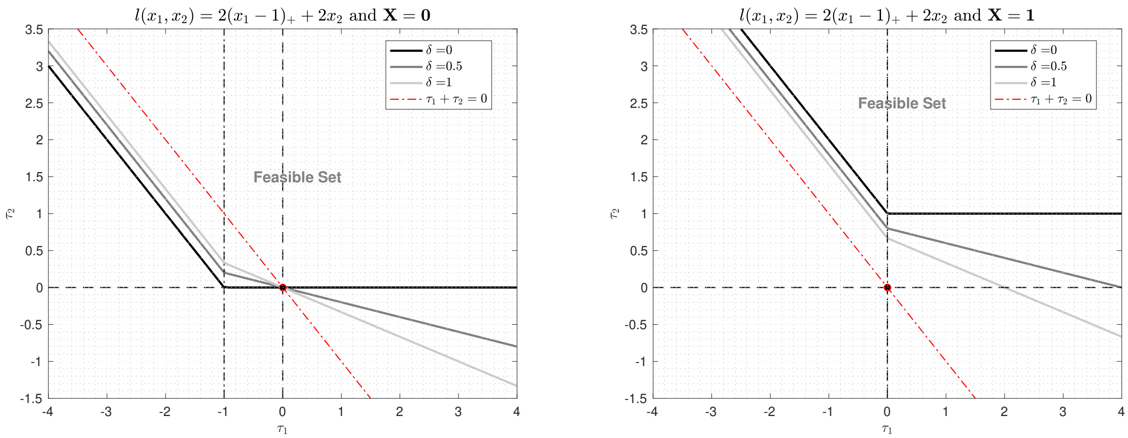
$$\rho_{l_1}^\delta(\mathbf{0}) = \inf_{\tau_1,\tau_2\in\mathbb{R}} \{\tau_1 + \tau_2 : 2(-\tau_1 - 1)_+ - 2\tau_2 \leq \delta(\tau_1 + \tau_2)\} = -\frac{2}{2+\delta}$$

and

$$\rho_{l_1}^\delta(\mathbf{1}) = \inf_{\tau_1,\tau_2\in\mathbb{R}} \{\tau_1 + \tau_2 : 2(-\tau_1)_+ - 2\tau_2 + 2 \leq \delta(\tau_1 + \tau_2)\} = \frac{2}{2+\delta}.$$

Note that in this case, the optimal values are bounded but the optimal solution sets are not. Likewise, for $l_2(x_1, x_2) = (x_1 - 1)_+ + 2x_2$, we can obtain

$$\rho_{l_2}^\delta(\mathbf{0}) = \inf_{\tau_1,\tau_2\in\mathbb{R}} \{\tau_1 + \tau_2 : (-\tau_1 - 1)_+ - 2\tau_2 \leq \delta(\tau_1 + \tau_2)\} = -\infty$$



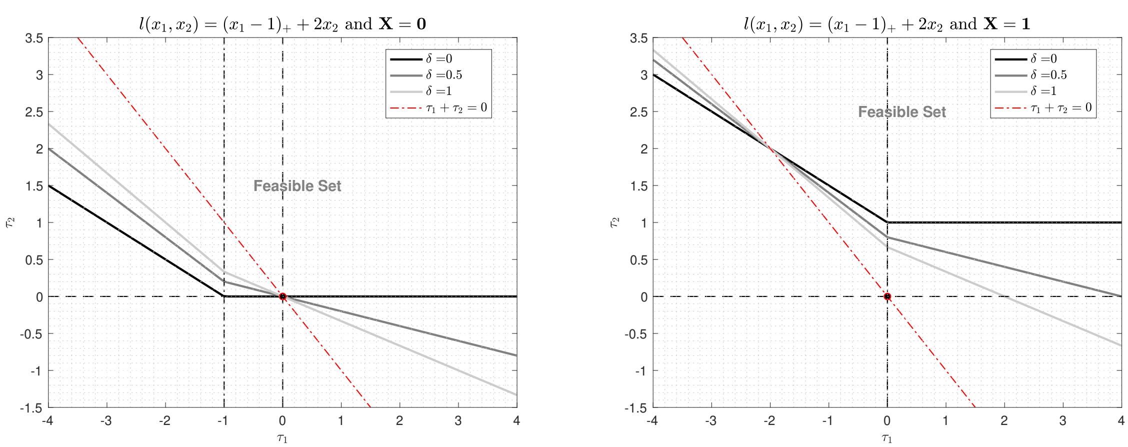
(a) **(b)**

Figure A.1: Feasible set to MLRM problem with $\delta = 0, 0.5, 1$ and $l = l_1$ for **(a)** $\boldsymbol{X} = \mathbf{0}$ and the dark/grey solid piecewise linear curves represent $2(-\tau_1 - 1)_+ - 2\tau_2 = \delta(\tau_1 + \tau_2)$ for $\delta = 0, 0.5, 1$, respectively, and **(b)** $\boldsymbol{X} = \mathbf{1}$ and the dark/grey solid piecewise linear curves represent $2(-\tau_1)_+ - 2\tau_2 + 2 = \delta(\tau_1 + \tau_2)$ for $\delta = 0, 0.5, 1$, respectively. Minimum is attained when the red dashed curve is shifted in parallel (left downward in (a) and right upward in (b)) to meet the dark solid curve at the boundary of the feasible sets. In both cases, there exist infinitely many optimal allocations to minimize the overall capital requirement.

and

$$\rho^{\delta}_{l_2}(\mathbf{1}) = \inf_{\tau_1,\tau_2 \in \mathbb{R}} \{\tau_1 + \tau_2 : (-\tau_1)_+ - 2\tau_2 + 2 \le \delta(\tau_1 + \tau_2)\} = -\infty$$

based on the feasible sets plotted in Figure A.2 (a) and (b), respectively.



(a) **(b)**

Figure A.2: Feasible set to MLRM problem with $\delta = 0, 0.5, 1$ and $l = l_2$ for **(a)** $\boldsymbol{X} = \mathbf{0}$ and the dark/grey solid piecewise linear curves represent $(-\tau_1 - 1)_+ - 2\tau_2 = \delta(\tau_1 + \tau_2)$ for $\delta = 0, 0.5, 1$, respectively, and **(b)** $\boldsymbol{X} = \mathbf{1}$ and the dark/grey solid piecewise linear curves represent $(-\tau_1)_+ - 2\tau_2 + 2 = \delta(\tau_1 + \tau_2)$ for $\delta = 0, 0.5, 1$, respectively. In both cases, the red dashed curve representing the contour of the objective function can be shifted left downwards infinitely with non-empty intersection with the feasible sets, which means the optimal value is $-\infty$.

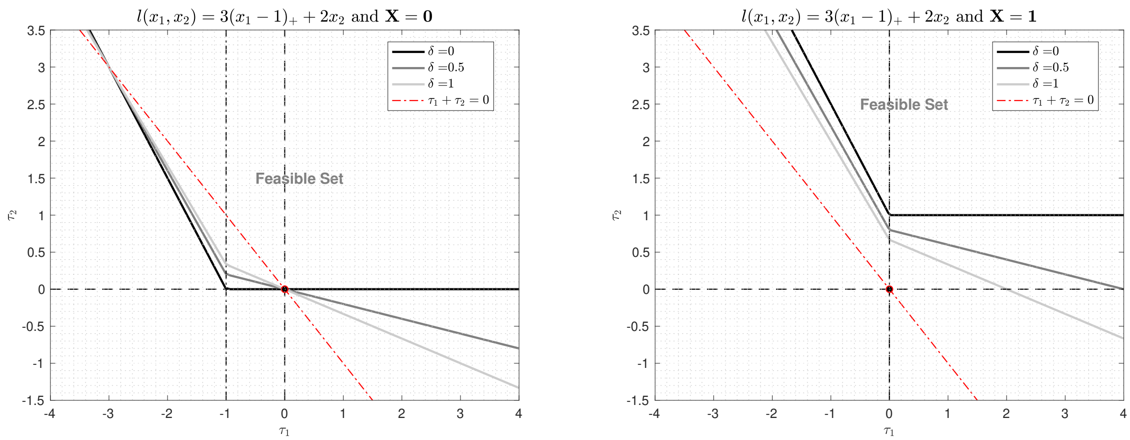
Consider another normalized multivariate loss function $l_3(x_1, x_2) = 3(x_1 - 1)_+ + 2x_2$. We can obtain

$$\rho^{\delta}_{l_3}(\mathbf{0}) = \inf_{\tau_1,\tau_2 \in \mathbb{R}} \{\tau_1 + \tau_2 : 3(-\tau_1 - 1)_+ - 2\tau_2 \le \delta(\tau_1 + \tau_2)\} = -\frac{2}{2+\delta}$$

and

$$\rho^{\delta}_{l_3}(\mathbf{1}) = \inf_{\tau_1,\tau_2\in\mathbb{R}} \{\tau_1 + \tau_2 : 3(-\tau_1)_+ - 2\tau_2 + 2 \le \delta(\tau_1 + \tau_2)\} = \frac{2}{2+\delta}$$

based on the feasible sets plotted in Figure A.3 (a) and (b), respectively. From Figures A.1-A.3, it is



(a) **(b)**

Figure A.3: Feasible set to MLRM problem with $\delta = 0, 0.5, 1$ and $l = l_3$ for **(a)** $\boldsymbol{X} = \mathbf{0}$ and the dark/grey solid piecewise linear curves represent $3(\tau_1 - 1)_+ - 2\tau_2 = \delta(\tau_1 + \tau_2)$ for $\delta = 0, 0.5, 1$, respectively, and **(b)** $\boldsymbol{X} = \mathbf{1}$, and the dark/grey solid piecewise linear curves represent $3(\tau_1)_+ - 2\tau_2 + 2 = \delta(\tau_1 + \tau_2)$ for $\delta = 0, 0.5, 1$, respectively. In both cases, there exists only one optimal allocation strategy to minimize the overall capital requirement.

evident that the feasible set for problem (2.1) is an unbounded convex set, which encompasses the upper set with the ordering cone $\mathbb{R}^n_+$.

From Example A.3, we can see that the MLRM heavily depends on the structure of the multivariate loss function. A small change of the structure of l may lead to huge change of the value of MLRM to negative infinite. In the next subsubsection, we will specify a class of l and a set of random vector $\boldsymbol{X}$ to ensure the proposed measure is finite-valued.

A.3 Multivariate Orlicz space

A function $\theta : \mathbb{R}^n \mapsto [0, \infty]$ is called a *multivariate Young function* if it is convex, lower semicontinuous with $\theta(\boldsymbol{x}) = \theta(|\boldsymbol{x}|)$, non-decreasing on $\mathbb{R}^n_+$, and satisfies

$$\lim_{\|\boldsymbol{x}\|\downarrow 0} \theta(\boldsymbol{x}) = \theta(\mathbf{0}) = 0 \text{ and } \lim_{\|\boldsymbol{x}\|\uparrow\infty} \theta(\boldsymbol{x}) = \infty.$$

The second equality means that $\theta(\boldsymbol{x})$ is coercive in $\mathbb{R}^n$. In particular, θ attains its minimum at $\mathbf{0}$. Its convex conjugate can be defined as

$$\theta^*(\boldsymbol{y}) := \sup_{\boldsymbol{x}\in\mathbb{R}^n} \{\boldsymbol{x}^T\boldsymbol{y} - \theta(\boldsymbol{x})\},\ \boldsymbol{y} \in \mathbb{R}^n,$$

which is again a multivariate Young function and its convex conjugate is θ, see Armenti et al. (2018, Lemma B.1). For any $\boldsymbol{X} \in L^0_n$, the Luxemburg norm of $\boldsymbol{X}$ is defined as

$$\|\boldsymbol{X}\|_\theta = \inf_{\lambda>0} \{\lambda : \mathbb{E}[\theta(\boldsymbol{X}/\lambda)] \le 1\},$$

where, by convention, $\inf\{t : t \in \emptyset\} = \infty$. The multivariate Orlicz space and multivariate Orlicz heart are respectively defined as

$$L_n^\theta := \{\boldsymbol{X} \in L_n^0 : \|\boldsymbol{X}\|_\theta < \infty\} = \{\boldsymbol{X} \in L_n^0 : \mathbb{E}[\theta(\lambda \boldsymbol{X})] < \infty \text{ for some } \lambda > 0\}$$

and

$$M_n^\theta := \{\boldsymbol{X} \in L_n^0 : \mathbb{E}[\theta(\lambda \boldsymbol{X})] < \infty \text{ for all } \lambda > 0\}.$$

We now summarize some key results about the multivariate Orlicz space in the next lemma based on Armenti et al. (2018, Lemma B.2 and Thetorem B.3).

Lemma A.1 *Let $\theta : \mathbb{R}^n \mapsto [0, \infty]$ be a multivariate Young function. Then the following assertions hold.*

(i) If $0 < \|\boldsymbol{X}\|_\theta < \infty$, then $\mathbb{E}[\theta(|\boldsymbol{X}|/\|\boldsymbol{X}\|_\theta)] \leq 1$. In particular, $\mathcal{B} := \{\boldsymbol{X} \in L_n^0 : \|\boldsymbol{X}\|_\theta \leq 1\} = \{\boldsymbol{X} \in L_n^0 : \mathbb{E}[\theta(|\boldsymbol{X}|)] \leq 1\}$.

(ii) The gauge function $\|\cdot\|_\theta : L_n^0 \mapsto \mathbb{R} \cup \{\infty\}$ is a norm both on the multivariate Orlicz space L_n^θ and on the multivariate Orlicz heart M_n^θ. Moreover, the normed spaces $(L_n^\theta, \|\cdot\|_\theta)$ and $(M_n^\theta, \|\cdot\|_\theta)$ are Banach spaces.

(iii) The Hölder inequality holds: $\mathbb{E}[|\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}|] \leq \|\boldsymbol{X}\|_\theta \|\boldsymbol{Y}\|_{\theta^*}$.

(iv) If θ is finite, then the topological dual of M_n^θ is $L_n^{\theta^*}$.

A.4 An illustrative figure

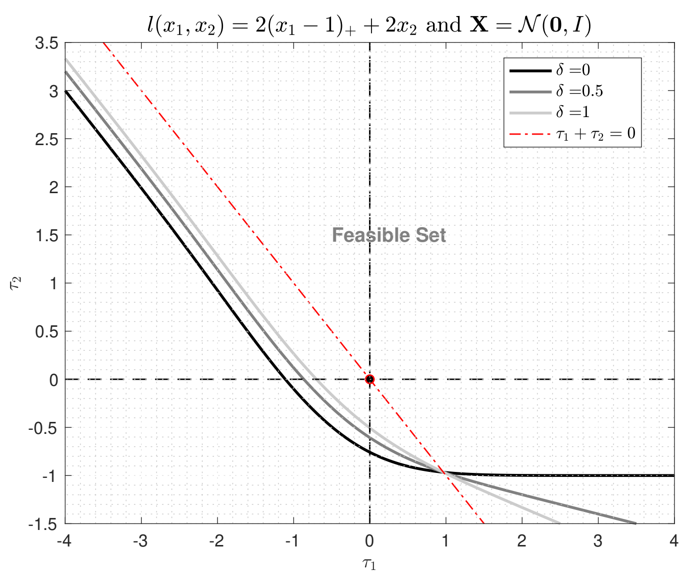


Figure A.4: Illustration of the feasible set to MLRM with $\delta = 0, 0.5, 1$ and $l = l_2$ for $\boldsymbol{X} = \mathcal{N}(\mathbf{0}, I)$.

A.5 Some classical facts in convex optimization

For an extended real-valued function f on a locally convex topological vector space X, its convex conjugate is defined as

$$f^*(x^*) = \sup_{x \in X}\{\langle x^*, x\rangle - f(x)\}, \ x^* \in X^*,$$

where X^* is the topological dual of X. The Fenchel-Moreau theorem states that if f is lower semicontinuous, convex, and proper, then so is f^*, and it holds

$$f(x) = f^{**}(x) = \sup_{x^* \in X^*} \{\langle x, x^* \rangle - f^*(x^*)\},\ x \in X.$$

Following Rockafellar (1997), for any nonempty set $C \subseteq \mathbb{R}^n$, we define its *recession cone* as

$$0^+ C := \{y \in \mathbb{R}^n : x + \lambda y \in C \text{ for every } x \in C \text{ and } \lambda \geq 0\}.$$

By Rockafellar (1997, Theorem 8.3), if C is nonempty, closed and convex, then

$$0^+ C = \{y \in \mathbb{R}^n : \text{ there exists } x \in C \text{ such that } x + \lambda y \in C \text{ for every } \lambda \geq 0\}.$$

Moreover, by Rockafellar (1997, Theorem 8.4), a nonempty, closed and convex set C is compact if and only if $0^+ C = \{0\}$.

We turn now to the application of the above results to convex functions. Given a proper, convex, and lower semicontinuous function f on $\mathbb{R}^n$, we call $y \in \mathbb{R}^n$ a *direction of recession* of f if there exists $x \in \mathrm{dom} f$ such that the map $\lambda \mapsto f(x + \lambda y)$ is decreasing on $\mathbb{R}_+$. We denote by $f0^+$ the *recession function* of f, that is, the function with epigraph given as the recession cone of the epigraph of f, and we call

$$0^+ f := \left\{y \in \mathbb{R}^n : (f0^+)(y) \leq 0\right\}$$

the *recession cone* of f. The following results gathers results from Rockafellar (1997, Pages 66-70).

Theorem A.1 *Let f be a proper, closed, and convex function on $\mathbb{R}^n$. Then the following holds:*

(i) Given $x, y \in \mathbb{R}^n$, if

$$\liminf_{\lambda \to \infty} f(x + \lambda y) < \infty,$$

then the mapping $\lambda \mapsto f(x + \lambda y)$ is a non-increasing function.

(ii) All the nonempty level sets of the form $B = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$, $\alpha \in \mathbb{R}$ have the same recession cone, namely the recession cone of f. That is,

$$0^+ f = 0^+ B \text{ for every } \alpha \in \mathbb{R} \text{ such that } B \neq \emptyset.$$

(iii) $f0^+$ is a positively homogeneous, proper, closed, and convex function, such that

$$f0^+(y) = \sup_{\lambda > 0} \frac{f(x + \lambda y) - f(x)}{\lambda} = \lim_{\lambda \to \infty} \frac{f(x + \lambda y) - f(x)}{\lambda},\ y \in \mathbb{R}^n,$$

for any $x \in \mathrm{dom} f$.

(iv) There exists $x \in \mathrm{dom} f$ such that the map $\lambda \mapsto f(x + \lambda y)$ is non-increasing on $\mathbb{R}_+$, that is, y is a direction of recession of f if and only if this map is decreasing for every $x \in \mathrm{dom} f$, which in turn is equivalent to $(f0^+)(y) \leq 0$.

(v) The map $\lambda \mapsto f(x + \lambda y)$ is constant on $\mathbb{R}_+$ for every $x \in \mathrm{dom} f$ if and only if $(f0^+)(y) \leq 0$ and $(f0^+)(-y) \leq 0$.

Appendix B Proofs of Statements

B.1 Proof of Proposition 2.1

Proof. This is directly from the properties of l and the linearity of the expectation operator. ■

B.2 Proof of Proposition 3.1

Proof. Since l satisfies Assumption 3.1, then one has (3.1). Consequently, for any $\boldsymbol{X} \in L_n^1$, $\rho_l^\delta(\boldsymbol{X}) > -\infty$. Since $L_n^\theta \subset L_n^1$, then it suffices to show that $\rho_l^\delta(\boldsymbol{X}) < \infty$ for $\boldsymbol{X} \in L_n^\theta$ and $\rho_l^\delta(\boldsymbol{X}) = \infty$ for $\boldsymbol{X} \notin L_n^\theta$.

Let $\boldsymbol{X} \in L_n^\theta$. By definition $\mathbb{E}[\theta(\lambda_0|\boldsymbol{X}|)] < \infty$ for some $\lambda_0 > 0$, i.e., $\mathbb{E}[l(\lambda_0|\boldsymbol{X}|)] < \infty$ for some $\lambda_0 > 0$. If $\lambda_0 > 1$, then by the convexity of l, one has

$$l(\boldsymbol{x} - \boldsymbol{\tau}) = l\left(\frac{1}{\lambda_0}(\lambda_0 \boldsymbol{x}) + \frac{\lambda_0 - 1}{\lambda_0}\left(-\frac{\lambda_0}{\lambda_0 - 1}\boldsymbol{\tau}\right)\right) \leq \frac{1}{\lambda_0} l(\lambda_0 \boldsymbol{x}) + \frac{\lambda_0 - 1}{\lambda_0} l\left(-\frac{\lambda_0}{\lambda_0 - 1}\boldsymbol{\tau}\right).$$

Consequently, it is straightforward that

$$\left\{\boldsymbol{\tau} \in \mathbb{R}^n : \frac{1}{\lambda_0}\mathbb{E}[l(\lambda_0 \boldsymbol{X})] + \frac{\lambda_0 - 1}{\lambda_0} l\left(-\frac{\lambda_0}{\lambda_0 - 1}\boldsymbol{\tau}\right) \leq \delta \sum_{i=1}^n \tau_i\right\} \subseteq \left\{\boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^N \tau_i\right\}.$$

Thus, by the definition of $\rho_l^\delta(\cdot)$, one has

$$\rho_l^\delta(\boldsymbol{X}) \leq \inf_{\boldsymbol{\tau} \in \mathbb{R}^n} \left\{\sum_{i=1}^n \tau_i : \frac{1}{\lambda_0}\mathbb{E}[l(\lambda_0 \boldsymbol{X})] + \frac{\lambda_0 - 1}{\lambda_0} l\left(-\frac{\lambda_0}{\lambda_0 - 1}\boldsymbol{\tau}\right) \leq \delta \sum_{i=1}^n \tau_i\right\} < \infty.$$

This is due to the fact that $\mathbb{E}[l(\lambda_0 \boldsymbol{X})] \leq \mathbb{E}[l(\lambda_0|\boldsymbol{X}|)] < \infty$ since l is non-decreasing and the mapping

$$\boldsymbol{\tau} \mapsto \psi(\boldsymbol{\tau}) := \delta \sum_{i=1}^n \tau_i - \frac{\lambda_0 - 1}{\lambda_0} l\left(-\frac{\lambda_0}{\lambda_0 - 1}\boldsymbol{\tau}\right)$$

is strictly increasing and unbounded from above and below, which means that there exists a $\widetilde{\boldsymbol{\tau}} \in \mathbb{R}^n$ such that $\mathbb{E}[l(\lambda_0 \boldsymbol{X})]/\lambda_0 \leq \psi(\widetilde{\boldsymbol{\tau}})$. Similar results can be obtained for the case when $\lambda_0 < 1$. When $\lambda_0 = 1$, one has $\mathbb{E}[l(|\boldsymbol{X}|)] < \infty$. Consequently, there exists a $\widetilde{\boldsymbol{\tau}} \in \mathbb{R}^n$ such that

$$\mathbb{E}[l(\boldsymbol{X} - \widetilde{\boldsymbol{\tau}})] \leq \delta \sum_{i=1}^n \widetilde{\tau}_i$$

since l is non-decreasing and $\mathbb{E}[l(\boldsymbol{X})] < \infty$.

Next, let $\boldsymbol{X} \notin L_n^\theta$. By definition $\mathbb{E}[l(\theta(\lambda|\boldsymbol{X}|))] = \infty$ for all $\lambda > 0$, i.e., $\mathbb{E}[l(\lambda|\boldsymbol{X}|)] = \infty$ for all $\lambda > 0$. Suppose for the sake of a contradiction that $\rho_l^\delta(\boldsymbol{X}) < \infty$. Then there exists a $\widetilde{\boldsymbol{\tau}} \in \mathbb{R}^n$ such that

$$\mathbb{E}[l(\boldsymbol{X} - \widetilde{\boldsymbol{\tau}})] \leq \delta \sum_{i=1}^n \widetilde{\tau}_i < \infty.$$

Moreover, such $\widetilde{\boldsymbol{\tau}}$ can be chosen as $\widehat{\tau}\mathbf{1} \in \mathbb{R}^n$ with $\widehat{\tau} = \max\{\widetilde{\tau}_i, i = 1, \ldots, n\}$. When $\widehat{\tau} \leq 0$, by the non-decreasing property of l, one has

$$\mathbb{E}[l(\boldsymbol{X})] \leq \mathbb{E}[l(\boldsymbol{X} - \widehat{\tau}\mathbf{1})] \leq n\delta\widehat{\tau} < \infty,$$

which contradicts the assumption that $\mathbb{E}[l(|\boldsymbol{X}|)] = \infty$. When $\widehat{\tau} > 0$, it follows from the convexity of l and $l(\mathbf{1}) = n$ that

$$\mathbb{E}[l(\alpha\boldsymbol{X} - \alpha\widehat{\tau}\mathbf{1} + (1-\alpha)\mathbf{1})] \leq \alpha\mathbb{E}[l(\boldsymbol{X} - \widehat{\tau}\mathbf{1})] + (1-\alpha)l(\mathbf{1}) \leq \alpha\mathbb{E}[l(\boldsymbol{X} - \widehat{\tau}\mathbf{1})] + (1-\alpha)n < \infty$$

for any $\alpha \in [0,1]$. In particular, one can choose $\widehat{\alpha} = 1/(1+\widehat{\tau}) \in [0,1]$ and consequently, $\mathbb{E}[l(\widehat{\alpha}\boldsymbol{X})] < \infty$, a contradiction. Therefore, we have $\rho_l^\delta(\boldsymbol{X}) = \infty$. This completes the proof. ∎

B.3 Proof of Theorem 4.1

Proof. The proofs are standard and we include a proof for the completeness of the paper.

Part (i). Monotonicity: Let $\boldsymbol{X}, \boldsymbol{Y} \in L_n^\theta$ and $\boldsymbol{X}(\omega) \leq \boldsymbol{Y}(\omega)$ for all $\omega \in \Omega$. Then for any $\boldsymbol{\tau} \in \mathbb{R}^n$, one has $\boldsymbol{X} - \boldsymbol{\tau} \leq \boldsymbol{Y} - \boldsymbol{\tau}$. Since l is nondecreasing, it follows that $\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \mathbb{E}[l(\boldsymbol{Y} - \boldsymbol{\tau})]$. Consequently,

$$\left\{\boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{Y} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^n \tau_i\right\} \subseteq \left\{\boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^n \tau_i\right\},$$

which implies $\rho_l^\delta(\boldsymbol{X}) \leq \rho_l^\delta(\boldsymbol{Y})$.

Convexity: Let $\widehat{\boldsymbol{\tau}} \in \mathcal{F}^\delta(\boldsymbol{X})$ and $\widetilde{\boldsymbol{\tau}} \in \mathcal{F}^\delta(\boldsymbol{Y})$ be ϵ optimal solutions of (2.1), that is,

$$\rho_l^\delta(\boldsymbol{X}) = \sum_{i=1}^n \hat{\tau}_i - \epsilon \quad \text{and} \quad \rho_l^\delta(\boldsymbol{Y}) = \sum_{i=1}^n \tilde{\tau}_i - \epsilon.$$

For any $\alpha \in [0,1]$, let $\boldsymbol{Z}_\alpha := \alpha\boldsymbol{X} + (1-\alpha)\boldsymbol{Y}$ and $\boldsymbol{\tau}_\alpha := (\alpha\widehat{\boldsymbol{\tau}} + (1-\alpha)\widetilde{\boldsymbol{\tau}})$. The convexity of l ensures that

$$\mathbb{E}[l(\boldsymbol{Z}_\alpha - \boldsymbol{\tau}_\alpha)] \leq \alpha\mathbb{E}[l(\boldsymbol{X} - \widehat{\boldsymbol{\tau}})] + (1-\alpha)\mathbb{E}[l(\boldsymbol{Y} - \widetilde{\boldsymbol{\tau}})] \leq \delta\left(\alpha\sum_{i=1}^n \widehat{\tau}_i + (1-\alpha)\sum_{i=1}^n \widetilde{\tau}_i\right),$$

which implies $\boldsymbol{\tau}_\alpha \in \mathcal{F}^\delta(\boldsymbol{Z}_\alpha)$. Consequently, we have

$$\rho_l^\delta(\boldsymbol{Z}_\alpha) \leq \alpha\sum_{i=1}^n \widehat{\tau}_i + (1-\alpha)\sum_{i=1}^n \widetilde{\tau}_i = \alpha\rho_l^\delta(\boldsymbol{X}) + (1-\alpha)\rho_l^\delta(\boldsymbol{Y}) + \epsilon.$$

The conclusion follows since $\epsilon > 0$ can be arbitrarily small.

Part (ii). Define the Lagrangian function of problem (2.1)

$$L(\boldsymbol{\tau}, \lambda; \boldsymbol{X}) := \sum_{i=1}^n \tau_i + \lambda\left[\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] - \delta\sum_{i=1}^n \tau_i\right] = (1-\lambda\delta)\sum_{i=1}^n \tau_i + \lambda\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})],$$

where $\lambda \geq 0$ is the Lagrange multiplier. Let

$$v(\lambda; \boldsymbol{X}) := \inf_{\boldsymbol{\tau} \in \mathbb{R}^n} L(\boldsymbol{\tau}, \lambda; \boldsymbol{X}) = \inf_{\boldsymbol{\tau} \in \mathbb{R}^n}\left\{(1-\lambda\delta)\sum_{i=1}^n \tau_i + \lambda\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})]\right\}.$$

Note that the mapping $\boldsymbol{X} \mapsto v(\lambda; \boldsymbol{X})$ is closely related to a multivariate version of the *optimized certainty equivalent* in Ben-Tal and Teboulle (2007). To see this, it suffices to show that $1 - \lambda\delta > 0$. If $1 - \lambda\delta < 0$, that is, $\lambda > 1/\delta$, then $v(\lambda; \boldsymbol{X}) = -\infty$. This is because $\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \mathbb{E}[l(\boldsymbol{X})]$ as $\boldsymbol{\tau}$ goes to positive infinity componentwise, and consequently $L(\boldsymbol{\tau}, \lambda; \boldsymbol{X}) \to -\infty$. If $1 - \lambda\delta = 0$, that is, $\lambda = 1/\delta > 0$, then $v(1/\delta; \boldsymbol{X}) = 1/\delta \cdot \inf_{\boldsymbol{\tau} \in \mathbb{R}^n} \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] = 1/\delta \cdot l(-\infty \cdot \mathbf{1}) \leq 0$.

Next, we focus on the case that $1-\lambda\delta>0$. Observe that

$$\rho_l^\delta(\boldsymbol{X}) = \inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\sup_{\lambda\geq 0} L(\boldsymbol{\tau},\lambda;\boldsymbol{X}) \geq \sup_{\lambda\geq 0}\inf_{\boldsymbol{\tau}\in\mathbb{R}^n} L(\boldsymbol{\tau},\lambda;\boldsymbol{X}) = \sup_{\lambda\geq 0} v(\lambda;\boldsymbol{X}).$$

Since $\delta>0$ and l is non-decreasing, then the mapping $\boldsymbol{\tau}\mapsto\delta\sum_{i=1}^n\tau_i-\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]$ is strictly increasing and unbounded from above. Consequently, there exists a $\widehat{\boldsymbol{\tau}}\in\mathbb{R}^n$ lying in the interior of the feasible set to problem (2.1) and so the Slater condition is fulfilled. Thus, there is no duality gap, see, e.g., Rockafellar (1997, Theorem 28.2). Namely, $\rho_l^\delta(\boldsymbol{X})=\sup_{\lambda\geq 0}v(\lambda;\boldsymbol{X})$. Note that $v(0;\boldsymbol{X})=\inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\sum_{i=1}^n\tau_i=-\infty$. Thus, we have

$$\rho_l^\delta(\boldsymbol{X}) = \sup_{\lambda>0} v(\lambda;\boldsymbol{X}).$$

Consequently, an easy multivariate adaptation of Ben-Tal and Teboulle (2007, Theorem 4.2) yields

$$v(\lambda;\boldsymbol{X}) = (1-\lambda\delta)\cdot\sup_{\boldsymbol{Q}\in\mathcal{Q}}\left\{\sum_{i=1}^n\mathbb{E}_{Q_i}[X_i]-\mathbb{E}\left[(l_{\lambda,\delta})^*\left(\frac{d\boldsymbol{Q}}{d\mathbb{P}}\right)\right]\right\}$$

where $l_{\lambda,\delta}(\boldsymbol{x})=\lambda/(1-\lambda\delta)l(\boldsymbol{x})$, hence $(l_{\lambda,\delta})^*(\boldsymbol{x})=\lambda/(1-\lambda\delta)l^*((1-\lambda\delta)/\lambda\cdot\boldsymbol{x})$. To see this, note that in our context $\boldsymbol{X}$ represents the loss whereas in Ben-Tal and Teboulle (2007), X denotes the financial position. Moreover, $l(x)=-u(-x)$. Thus, $\sup_{\eta\in\mathbb{R}}\{\eta+\mathbb{E}[u(X-\eta)]\}=-\inf_{\eta\in\mathbb{R}}\{\eta+\mathbb{E}[l(-X-\eta)]\}=\inf_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[-X]+\mathbb{E}[l^*(dQ/d\mathbb{P})]\}$ and consequently, $\inf_{\eta\in\mathbb{R}}\{\eta+\mathbb{E}[l(-X-\eta)]\}=-\inf_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[-X]+\mathbb{E}[l^*(dQ/d\mathbb{P})]\}=\sup_{Q\in\mathcal{Q}}\{\mathbb{E}_Q[X]-\mathbb{E}[l^*(dQ/d\mathbb{P})]\}$. Therefore, one has

$$\begin{aligned}
\sup_{\lambda>0} v(\lambda;\boldsymbol{X}) &= \sup_{\lambda>0}\left\{(1-\lambda\delta)\cdot\sup_{\boldsymbol{Q}\in\mathcal{Q}}\left\{\sum_{i=1}^n\mathbb{E}_{Q_i}[X_i]-\mathbb{E}\left[(l_{\lambda,\delta})^*\left(\frac{d\boldsymbol{Q}}{d\mathbb{P}}\right)\right]\right\}\right\}\\
&= \sup_{\boldsymbol{Q}\in\mathcal{Q}}\sup_{\lambda>0}\left\{(1-\lambda\delta)\cdot\sum_{i=1}^n\mathbb{E}_{Q_i}[X_i]-(1-\lambda\delta)\cdot\mathbb{E}\left[(l_{\lambda,\delta})^*\left(\frac{d\boldsymbol{Q}}{d\mathbb{P}}\right)\right]\right\}\\
&= \sup_{\boldsymbol{Q}\in\mathcal{Q}}\left\{\sum_{i=1}^n\mathbb{E}_{Q_i}[X_i]-\inf_{\lambda>0}\left\{\lambda\mathbb{E}\left[l^*\left(\frac{1-\lambda\delta}{\lambda}\frac{d\boldsymbol{Q}}{d\mathbb{P}}\right)\right]+\lambda\delta\sum_{i=1}^n\mathbb{E}_{Q_i}[X_i]\right\}\right\}.
\end{aligned}$$

Part (iii). Let $\boldsymbol{X}\in L_n^\theta$ be fixed and $\boldsymbol{Y}\in L_n^\theta$ with $\boldsymbol{Y}\xrightarrow{a.s.}\boldsymbol{X}$. By Lemma 3.1, both $\rho_l^\delta(\boldsymbol{X})$ and $\rho_l^\delta(\boldsymbol{Y})$ are finite-valued. Let $\mathcal{F}^\delta(\cdot)$ be defined as in (2.5). Under our assumption, we can write $\rho_l^\delta(\boldsymbol{X})$ and $\rho_l^\delta(\boldsymbol{Y})$ respectively as

$$\rho_l^\delta(\boldsymbol{X}) = \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{X})}\left\{\sum_{i=1}^n\tau_i\right\} = \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{X})\cap B(\boldsymbol{0},\Delta)}\left\{\sum_{i=1}^n\tau_i\right\}$$

and

$$\rho_l^\delta(\boldsymbol{Y}) = \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{Y})}\left\{\sum_{i=1}^n\tau_i\right\} = \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\boldsymbol{0},\Delta)}\left\{\sum_{i=1}^n\tau_i\right\}$$

for all $\boldsymbol{Y}$ in a neighborhood of $\boldsymbol{X}$ equipped with Luxemburg norm $\|\cdot\|_\theta$. Consequently, we have

$$\begin{aligned}
\rho_l^\delta(\boldsymbol{Y})-\rho_l^\delta(\boldsymbol{X}) &= \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\boldsymbol{0},\Delta)}\sum_{i=1}^n\tau_i-\inf_{\widetilde{\boldsymbol{\tau}}\in\mathcal{F}^\delta(\boldsymbol{X})\cap B(\boldsymbol{0},\Delta)}\sum_{i=1}^n\widetilde{\tau}_i\\
&= \sup_{\widetilde{\boldsymbol{\tau}}\in\mathcal{F}^\delta(\boldsymbol{X})\cap B(\boldsymbol{0},\Delta)}\inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\boldsymbol{0},\Delta)}\left(\sum_{i=1}^n\tau_i-\sum_{i=1}^n\widetilde{\tau}_i\right)\\
&= \sup_{\widetilde{\boldsymbol{\tau}}\in\mathcal{F}^\delta(\boldsymbol{X})\cap B(\boldsymbol{0},\Delta)}\inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\boldsymbol{0},\Delta)}\sum_{i=1}^n(\tau_i-\widetilde{\tau}_i)
\end{aligned}$$

$$
\begin{aligned}
&\leq \sup_{\widetilde{\boldsymbol{\tau}}\in\mathcal{F}^\delta(\boldsymbol{X})\cap B(\mathbf{0},\Delta)} \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\mathbf{0},\Delta)} \|\boldsymbol{\tau}-\widetilde{\boldsymbol{\tau}}\|_1 \\
&= \mathbb{D}_1(\mathcal{F}^\delta(\boldsymbol{X})\cap B(\mathbf{0},\Delta), \mathcal{F}^\delta(\boldsymbol{Y})\cap B(\mathbf{0},\Delta)).
\end{aligned}
$$

Likewise, we can establish

$$
\rho_l^\delta(\boldsymbol{X}) - \rho_l^\delta(\boldsymbol{Y}) \leq \mathbb{D}_1(\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\mathbf{0},\Delta), \mathcal{F}^\delta(\boldsymbol{X})\cap B(\mathbf{0},\Delta)).
$$

Combining the above two inequalities, we obtain

$$
|\rho_l^\delta(\boldsymbol{X}) - \rho_l^\delta(\boldsymbol{Y})| \leq \mathbb{H}_1(\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\mathbf{0},\Delta), \mathcal{F}^\delta(\boldsymbol{X})\cap B(\mathbf{0},\Delta)),
$$

where $\mathbb{H}_1$ is the Hausdorff distance induced by the l_1-norm. Since $\boldsymbol{Y} \xrightarrow{a.s.} \boldsymbol{X}$, then by Proposition 2.1 and Ruszczyński and Shapiro (2006, Proposition 3.1), we have $\mathbb{E}[l(\boldsymbol{Y}-\boldsymbol{\tau})] - \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})] \to 0$ for any $\boldsymbol{\tau}\in\mathbb{R}^n$. Since the inequalities defining $\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\mathbf{0},\Delta)$ and $\mathcal{F}^\delta(\boldsymbol{X})\cap B(\mathbf{0},\Delta)$ are convex systems which satisfy Slater constraint qualification (by making Δ is chosen sufficiently large if necessarily), then we can use Robinson's theorem (Robinson, 1975) to assert that $\mathbb{H}_1(\mathcal{F}^\delta(\boldsymbol{Y})\cap B(\mathbf{0},\Delta), \mathcal{F}^\delta(\boldsymbol{X})\cap B(\mathbf{0},\Delta)) \to 0$. This completes the proof. ■

B.4 Proof of Proposition 4.1

Proof. For $\gamma > 1$, since l is convex and $l(\mathbf{0}) = 0$, then $\gamma l(\boldsymbol{x}) \leq l(\gamma\boldsymbol{x})$ for any $\boldsymbol{x}\in\mathbb{R}^n$. Consequently,

$$
\left\{\gamma\boldsymbol{\tau}\in\mathbb{R}^n : \mathbb{E}[l(\gamma\boldsymbol{X}-\gamma\boldsymbol{\tau})] \leq \delta\sum_{i=1}^n \gamma\tau_i\right\} \subseteq \left\{\gamma\boldsymbol{\tau}\in\mathbb{R}^n : \gamma\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})] \leq \gamma\cdot\delta\sum_{i=1}^n \tau_i\right\}.
$$

By the definition of $\rho_l^\delta(\boldsymbol{X})$, the inclusion above implies that $\rho_l^\delta(\gamma\boldsymbol{X}) \geq \gamma\rho_l^\delta(\boldsymbol{X})$. Likewise, for $0\leq\gamma\leq 1$, we have $\gamma l(\boldsymbol{x}) \geq l(\gamma\boldsymbol{x})$ and subsequently $\rho_l^\delta(\gamma\boldsymbol{X}) \leq \gamma\rho_l^\delta(\boldsymbol{X})$.

Let $\gamma_2 > \gamma_1 > 0$. By the convexity of l, we have

$$
\frac{l(\gamma_2\boldsymbol{x}) - l(\gamma_1\boldsymbol{x})}{\gamma_2-\gamma_1} \geq \frac{l(\gamma_1\boldsymbol{x}) - l(\mathbf{0})}{\gamma_1 - 0},\ \forall\, \boldsymbol{x}\in\mathbb{R}^n.
$$

Since $l(\mathbf{0}) = 0$, then we have $l_{\gamma_2}(\boldsymbol{x}) \geq l_{\gamma_1}(\boldsymbol{x})$ for any $\boldsymbol{x}\in\mathbb{R}^n$. Following a similar argument to the proof of superhomogeneity of ρ_l^δ, we can show that the MLRM corresponding to l_γ satisfies $\rho_{l_{\gamma_2}}^\delta(\boldsymbol{X}) \geq \rho_{l_{\gamma_1}}^\delta(\boldsymbol{X})$. The conclusion follows from (4.2). ■

B.5 Proof of Proposition 4.2

Proof. The assertion follows immediately from the fact that if l is a supermodular, directionally convex, or Δ-monotone function, so is $l(\cdot - \boldsymbol{\tau})$ for any $\boldsymbol{\tau}\in\mathbb{R}^n$, see Armenti et al. (2018). Therefore, for instance, if $\boldsymbol{X} \succeq^{sm} \boldsymbol{Y}$ according to l, it follows that $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})] \geq \mathbb{E}[l(\boldsymbol{Y}-\boldsymbol{\tau})]$, which means

$$
\left\{\boldsymbol{\tau}\in\mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})] \leq \delta\sum_{i=1}^n \tau_i\right\} \subseteq \left\{\boldsymbol{\tau}\in\mathbb{R}^n : \mathbb{E}[l(\boldsymbol{Y}-\boldsymbol{\tau})] \leq \delta\sum_{i=1}^n \tau_i\right\}.
$$

Thus, $\rho_l^\delta(\boldsymbol{X}) \geq \rho_l^\delta(\boldsymbol{Y})$. The other two cases can be proved similarly. ■

B.6 Proof of Theorem 5.1

Proof. Part (i). Since $\boldsymbol{X} \in L_n^{\theta}$, by Lemma 3.1, $\rho_l^{\delta}(\boldsymbol{X})$ is finite-valued. Moreover, by the lower semicontinuity and convexity of l, the feasible set to problem (2.1), i.e.,

$$\mathcal{F}^{\delta}(\boldsymbol{X}) := \left\{\boldsymbol{\tau} \in \mathbb{R}^n : \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \leq \delta \sum_{i=1}^{n} \tau_i \right\}$$

is a nonempty closed convex set in $\mathbb{R}^n$. Thus, there exists a $\widetilde{\boldsymbol{\tau}} \in \mathbb{R}^n$ such that the level set

$$\mathcal{F}^{\delta}(\boldsymbol{X}) \bigcap \left\{\boldsymbol{\tau} \in \mathbb{R}^n : \sum_{i=1}^{n} \tau_i \leq \sum_{i=1}^{n} \widetilde{\tau}_i \right\} \neq \emptyset.$$

Since l is non-decreasing along each direction $d \in \mathbb{R}^n_+$, $\boldsymbol{\tau} \in \mathcal{F}^{\delta}(\boldsymbol{X})$ for all $\boldsymbol{\tau} \in \mathbb{R}^n_+$ with $\sum_{i=1}^n \tau_i$ being sufficiently large, then the feasible set $\mathcal{F}^{\delta}(\boldsymbol{X})$ is unbounded. On the other hand, $\mathcal{F}^{\delta}(\boldsymbol{X}) \neq \mathbb{R}^n$ because $\boldsymbol{\tau} \notin \mathcal{F}^{\delta}(\boldsymbol{X})$ for all $\boldsymbol{\tau} \in \mathbb{R}^n_-$ with $\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau})] \geq 0$. This means that we can find a $\widehat{\boldsymbol{\tau}} \in \mathbb{R}^n$ such that $\widehat{\boldsymbol{\tau}} \notin \mathcal{F}^{\delta}(X)$. Summarizing the discussions above, we have

$$\mathbb{E}[l(\boldsymbol{X} - \widetilde{\boldsymbol{\tau}})] \leq \delta \sum_{i=1}^{n} \widetilde{\tau}_i \quad \text{and} \quad \mathbb{E}[l(\boldsymbol{X} - \widehat{\boldsymbol{\tau}})] > \delta \sum_{i=1}^{n} \widehat{\tau}_i.$$

Thus, we can assert that the optimum is attained in the region

$$\widehat{A}(\boldsymbol{X}) := \mathcal{F}^{\delta}(\boldsymbol{X}) \bigcap \left\{\boldsymbol{\tau} \in \mathbb{R}^n : \sum_{i=1}^{n} \hat{\tau}_i \leq \sum_{i=1}^{n} \tau_i \leq \sum_{i=1}^{n} \widetilde{\tau}_i \right\}.$$

Under Assumption 5.1, $\hat{A}(\boldsymbol{X})$ is a compact set. Consequently, we can reformulate problem (2.1) as

$$\rho_l^{\delta}(\boldsymbol{X}) = \inf \left\{ \sum_{i=1}^{n} \tau_i : \boldsymbol{\tau} \in \widehat{A}(\boldsymbol{X}) \right\}. \tag{B.1}$$

Existence of an optimal solution follows by Weierstrass theorem. The convexity of the set of optimal solutions follows from the fact that this is a convex program. To see that the optimium is attained in set $\mathcal{F}^{\delta}_{=}(\boldsymbol{X})$ as defined in (2.6), we can show by contradiction that if it attains some $\boldsymbol{\tau}_0 \in \widehat{A}(\boldsymbol{X}) \backslash \mathcal{F}^{\delta}_{=}(\boldsymbol{X})$, then we can always find a feasible solution near $\boldsymbol{\tau}_0$ such that the objective function at the latter is smaller, a contradiction.

Part (ii). As we explained in the proof of Theorem 4.1 (ii), problem (2.1) satisfies the Slater constraint qualification. By the standard results of Lagrange multipliers for convex programs (see, e.g., Rockafellar (1997, Section 28)), we assert that the Lagrangian $L(\boldsymbol{\tau}, \lambda; \boldsymbol{X})$ of problem (2.1) has a saddle-point, denoted by $(\boldsymbol{\tau}^*, \lambda^*)$, iff $(\boldsymbol{\tau}^*, \lambda^*)$ satisfies the following Karush-Kuhn-Tucker (KKT) condition:

$$\lambda^* > 0,\ \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau}^*)] = \delta \sum_{i=1}^{n} \tau_i^* \text{ and } \mathbf{0} \in \partial_{\boldsymbol{\tau}} L(\boldsymbol{\tau}^*, \lambda^*; \boldsymbol{X}).$$

Since $\partial_{\boldsymbol{\tau}} L(\boldsymbol{\tau}^*, \lambda^*; \boldsymbol{X}) = (1 - \lambda\delta) \cdot \mathbf{1} - \lambda^* \mathbb{E}[\partial l(\boldsymbol{X} - \boldsymbol{\tau}^*)]$, then the optimal solutions to problem (2.1) are characterized by (5.3).

Part (iii). Finally, we show the uniqueness of the optimal solution under the strict convexity of l. By Part (i), the set of the optimal solutions lie in $\mathcal{F}^{\delta}_{=}(\boldsymbol{X})$. The strict convexity of l ensures that $\mathcal{F}^{\delta}_{=}(\boldsymbol{X}) \cap \{\boldsymbol{\tau} \in \mathbb{R}^n : \sum_{i=1}^n \tau_i = \rho_l^{\delta}(\boldsymbol{X})\}$ is a singleton. ■

B.7 Proof of Theorem 5.2

Proof. Part (i). Consider the Lagrangian function to problem (5.11)

$$\begin{aligned}\widetilde{L}(\boldsymbol{a},\eta;\boldsymbol{X}) &:= \mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{a})\right]+\frac{1}{\lambda^*}\sum_{i=1}^n a_i+\eta\left(\rho_l^\delta(\boldsymbol{X})-\sum_{i=1}^n a_i\right)\\ &= \mathbb{E}\left[l(\boldsymbol{X}-\boldsymbol{a})\right]+\left(\frac{1}{\lambda^*}-\eta\right)\sum_{i=1}^n a_i+\eta\rho_l^\delta(\boldsymbol{X}),\end{aligned}$$

where $\eta \geq 0$ is a parameter. Since the problem satisfies Slater condition, it follows by the discussions in Rockafellar (1997, Section 28)), a pair $(\boldsymbol{a}^*,\eta^*)$ is a saddle point of the Lagrangian $\widetilde{L}(\boldsymbol{a},\eta;\boldsymbol{X})$ of problem (5.11) iff $(\boldsymbol{a}^*,\eta^*)$ satisfies

$$\left(\frac{1}{\lambda^*}-\eta^*\right)\cdot\mathbf{1}\in\mathbb{E}[\partial l(\boldsymbol{X}-\boldsymbol{a}^*)] \quad \text{and} \quad \sum_{i=1}^n a_i=\rho_l^\delta(\boldsymbol{X}).$$

On the other hand, it follows from Theorem 5.1 that

$$(1-\lambda^*\delta)\cdot\mathbf{1}\in\lambda^*\mathbb{E}[\partial l(\boldsymbol{X}-\boldsymbol{\tau}^*)] \quad \text{and} \quad \sum_{i=1}^n \tau_i^*=\rho_l^\delta(\boldsymbol{X}). \tag{B.2}$$

This shows that $(\boldsymbol{\tau}^*,\delta)$ is a saddle point of $\widetilde{L}(\boldsymbol{a},\eta;\boldsymbol{X})$ and hence $\boldsymbol{\tau}^*$ is an optimal solution to problem (5.11).

Part (ii). The conclusion follows directly from Bonnans and Shapiro (2013, Theorem 4.25). ■

B.8 Proof of Theorem 6.1

Proof. Part (i). Let $\psi_\delta(\boldsymbol{\tau}) := \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})]-\delta\sum_{i=1}^n\tau_i$. Then problem (2.1) can be written as

$$\inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\left\{\sum_{i=1}^n\tau_i : 0\in\psi_\delta(\boldsymbol{\tau})+\mathbb{R}_+\right\}. \tag{B.3}$$

Since l is lower semicontinuous and convex in $\mathbb{R}^n$, then $\psi_\delta(\boldsymbol{\tau})$ is lower semicontinuous and convex in $\boldsymbol{\tau}$ and linear in δ. Since $\psi_\delta(\cdot)$ satisfies Slater constraint qualification, that is, there exists a point $\boldsymbol{\tau}^s$ and a real number $\eta>0$ such that $\psi_\delta(\boldsymbol{\tau}^s)\leq-\eta$, then by Robinson's error bound theorem for convex systems (Robinson, 1975),

$$\mathsf{dl}_1(\boldsymbol{\tau},\mathcal{F}^\delta(\boldsymbol{X}))\leq\frac{1}{\eta}\|\boldsymbol{\tau}-\boldsymbol{\tau}^s\|_1\cdot\max\{0,\psi_\delta(\boldsymbol{\tau})\} \text{ for any } \boldsymbol{\tau}\in\mathbb{R}^n, \tag{B.4}$$

where $\mathcal{F}^\delta(\boldsymbol{X}) := \{\boldsymbol{\tau}\in\mathbb{R}^n : \psi_\delta(\boldsymbol{\tau})\leq 0\}$ is the feasible set to problem (2.1). Let $\boldsymbol{\tau}^\delta$ be an optimal solution to problem (2.1). Then by Theorem 5.1, $\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau}^\delta)]=\delta\sum_{i=1}^n\tau_i^\delta$. Consequently, for any $\eta>0$, by the non-decreasing property of l, we have

$$\mathbb{E}[l(\boldsymbol{X}-(\boldsymbol{\tau}^\delta+\eta/\delta\boldsymbol{e}_k))]\leq\mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau}^\delta)]=\delta\sum_{i=1}^n\tau_i^\delta=\delta\left(\sum_{i=1}^n\tau_i^\delta+\frac{\eta}{\delta}\right)-\eta,$$

which implies that $\psi_\delta(\boldsymbol{\tau}^\delta+\eta/\delta\boldsymbol{e}_k)\leq-\eta$. Thus, we choose $\boldsymbol{\tau}^s=\boldsymbol{\tau}^\delta+\eta/\delta\boldsymbol{e}_k$ in the rest of the proof. Let $\boldsymbol{\tau}^{\delta'}\in\mathcal{S}^{\delta'}(\boldsymbol{X})\cap B(\mathbf{0},\Delta)\neq\emptyset$ be an optimal solution to problem (2.1) with risk tolerance level $\delta'>0$. By

Theorem 5.1, $\psi_{\delta'}(\boldsymbol{\tau}^{\delta'}) = 0$. Consequently, from (B.4), we have

$$\begin{aligned} \mathsf{dl}_1(\boldsymbol{\tau}^{\delta'}, \mathcal{F}^{\delta}(\boldsymbol{X})) &\leq \frac{1}{\eta}\|\boldsymbol{\tau}^{\delta'} - \boldsymbol{\tau}^{\delta} - \eta/\delta \boldsymbol{e}_k\|_1 \cdot \max\left\{0, \psi_{\delta}(\boldsymbol{\tau}^{\delta'})\right\} \\ &\leq \frac{1}{\eta}(\|\boldsymbol{\tau}^{\delta'} - \boldsymbol{\tau}^{\delta}\|_1 + \eta/\delta) \cdot \max\left\{0, \psi_{\delta}(\boldsymbol{\tau}^{\delta'}) - \psi_{\delta'}(\boldsymbol{\tau}^{\delta'})\right\} \\ &= \left(\frac{1}{\eta}\|\boldsymbol{\tau}^{\delta'} - \boldsymbol{\tau}^{\delta}\|_1 + \frac{1}{\delta}\right) \cdot \max\left\{0, (\delta' - \delta)\sum_{i=1}^{n} \tau_i^{\delta'}\right\}, \end{aligned}$$

Since $\eta > 0$ can be arbitrarily large, then

$$\mathsf{dl}_1(\boldsymbol{\tau}^{\delta'}, \mathcal{F}^{\delta}(\boldsymbol{X})) \leq \frac{1}{\delta}\max\left\{0, (\delta' - \delta)\sum_{i=1}^{n} \tau_i^{\delta'}\right\} = \frac{1}{\delta}\max\{0, (\delta' - \delta)\rho_l^{\delta'}(X)\}.$$

Analogous to the proof of Theorem 4.1 for the continuity, we have

$$\rho_l^{\delta}(\boldsymbol{X}) - \rho_l^{\delta'}(\boldsymbol{X}) = \inf_{\boldsymbol{\tau}\in\mathcal{F}^{\delta}(\boldsymbol{X})}\sum_{i=1}^{n}\tau_i - \sum_{i=1}^{n}\tau_i^{\delta'} \leq \inf_{\boldsymbol{\tau}\in\mathcal{F}^{\delta}(\boldsymbol{X})}\sum_{i=1}^{n}|\tau_i - \tau_i^{\delta'}| = \mathsf{dl}_1(\boldsymbol{\tau}^{\delta'}, \mathcal{F}^{\delta}(\boldsymbol{X})).$$

Combining the above two inequalities yields

$$\rho_l^{\delta}(\boldsymbol{X}) - \rho_l^{\delta'}(\boldsymbol{X}) \leq \frac{1}{\delta}\max\{0, (\delta' - \delta)\rho_l^{\delta'}(\boldsymbol{X})\}. \tag{B.5}$$

Likewise, we can establish

$$\rho_l^{\delta'}(\boldsymbol{X}) - \rho_l^{\delta}(\boldsymbol{X}) \leq \frac{1}{\delta'}\max\{0, (\delta - \delta')\rho_l^{\delta}(\boldsymbol{X})\}$$

and hence (6.1).

Part (ii). Under the additional conditions (a) and (b), we know by Theorem 5.1 that MLRM problem (2.1) has an unique solution for all tolerance parameter δ' close to δ, i.e., $\boldsymbol{\tau}^{\delta}$ and $\boldsymbol{\tau}^{\delta'}$ in (6.2) are well-defined. Moreover, since $\Lambda(\cdot)$ is a non-decreasing function, then its generalized inverse is well-defined. Thus

$$\Lambda(\|\boldsymbol{\tau}^{\delta'} - \boldsymbol{\tau}^{\delta}\|_1) \leq \sum_{i=1}^{n}\tau_i^{\delta'} - \sum_{i=1}^{n}\tau_i^{\delta} = \rho_l^{\delta'}(\boldsymbol{X}) - \rho_l^{\delta'}(\boldsymbol{X}).$$

Combining with the inequality (B.5), we obtain

$$\|\boldsymbol{\tau}^{\delta'} - \boldsymbol{\tau}^{\delta}\|_1 \leq \Lambda^{-1}(\rho_l^{\delta'}(\boldsymbol{X}) - \rho_l^{\delta'}(\boldsymbol{X})) \leq \Lambda^{-1}(\delta^{-1}\max\{0, (\delta' - \delta)\rho_l^{\delta'}(\boldsymbol{X})\})$$

which implies

$$\|\boldsymbol{\tau}^{\delta'} - \boldsymbol{\tau}^{\delta}\|_1 \leq \Lambda^{-1}(\delta^{-1}\max\{0, (\delta' - \delta)\rho_l^{\delta'}(\boldsymbol{X})\}).$$

Likewise, we can derive

$$\|\boldsymbol{\tau}^{\delta} - \boldsymbol{\tau}^{\delta'}\|_1 \leq \Lambda^{-1}(\delta'^{-1}\max\{0, (\delta - \delta')\rho_l^{\delta}(\boldsymbol{X})\}).$$

Combining the above inequalities, we obtain (6.2). ■

B.9 Proof of Theorem 6.2

Proof. Part (i). Analogous to the proof of Theorem 6.1, we have

$$
\begin{aligned}
\rho_l^\delta(\boldsymbol{X}) - \rho_l^\delta(\boldsymbol{Y}) &= \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{X})} \sum_{i=1}^n \tau_i - \sum_{i=1}^n \tau_i^{\boldsymbol{Y}} = \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{X})} \left(\sum_{i=1}^n \tau_i - \sum_{i=1}^n \tau_i^{\boldsymbol{Y}}\right) \\
&\leq \inf_{\boldsymbol{\tau}\in\mathcal{F}^\delta(\boldsymbol{X})} \sum_{i=1}^n |\tau_i - \tau_i^{\boldsymbol{Y}}| = \mathsf{dl}_1(\boldsymbol{\tau}^\delta(\boldsymbol{Y}), \mathcal{F}^\delta(\boldsymbol{X})) \\
&\leq \frac{1}{t}\|\boldsymbol{\tau}(\boldsymbol{Y}) - \boldsymbol{\tau}^s\|_1 \cdot \max\{0, \psi_\delta(\boldsymbol{\tau}(\boldsymbol{Y}))\},
\end{aligned}
$$

where $\psi_\delta(\boldsymbol{\tau}) := \mathbb{E}[l(\boldsymbol{X}-\boldsymbol{\tau})] - \delta\sum_{i=1}^n \tau_i$ and t is a positive constant satisfying $\psi_\delta(\boldsymbol{\tau}^\delta + \eta/\delta\boldsymbol{e}_k) \leq -t$ and $\boldsymbol{\tau}^s = \boldsymbol{\tau}^\delta + t/\delta\boldsymbol{e}_k$. Since t can be arbitrarily large, then following an argument in the proof of Theorem 6.1, we can show that

$$
\begin{aligned}
\rho_l^\delta(\boldsymbol{X}) - \rho_l^\delta(\boldsymbol{Y}) &\leq \frac{1}{\delta}\max\{0, \psi_\delta(\boldsymbol{\tau}(\boldsymbol{Y}))\} \\
&= \frac{1}{\delta}\max\{0, \psi_\delta(\boldsymbol{\tau}(\boldsymbol{Y})) - \psi_\delta(\boldsymbol{\tau}(\boldsymbol{X}))\} \\
&= \frac{1}{\delta}\max\left\{0, \mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau}^\delta(\boldsymbol{Y})) - l(\boldsymbol{Y} - \boldsymbol{\tau}^\delta(\boldsymbol{Y}))]\right\}.
\end{aligned}
$$

Since l is Lipschitz continuous with modulus L, then

$$
\left|\mathbb{E}[l(\boldsymbol{X} - \boldsymbol{\tau}^\delta(\boldsymbol{Y}))] - \mathbb{E}[l(\boldsymbol{Y} - \boldsymbol{\tau}^\delta(\boldsymbol{Y}))]\right| \leq L\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1].
$$

Combining the above two inequalities, we obtain

$$
\rho_l^\delta(\boldsymbol{X}) - \rho_l^\delta(\boldsymbol{Y}) \leq \frac{L}{\delta}\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1].
$$

Likewise, we can establish under condition (b)

$$
\rho_l^\delta(\boldsymbol{Y}) - \rho_l^\delta(\boldsymbol{X}) \leq \frac{L}{\delta}\mathbb{E}[\|\boldsymbol{Y} - \boldsymbol{X}\|_1].
$$

A combination of the two inequalities yields (6.3).

Part (ii). We show (6.4). Under the additional conditions (c) and (d), we know by Theorem 5.1 that MLRM problem (2.1) has an unique solution for all tolerance parameter δ' close to δ, i.e., $\boldsymbol{\tau}^\delta$ and $\boldsymbol{\tau}^{\delta'}$ in (6.2) are well-defined. Moreover, since $\Lambda(\cdot)$ is a non-decreasing function, then its generalized inverse is well-defined. Then for any $\boldsymbol{\tau}^{\boldsymbol{X}} \in \boldsymbol{\tau}^\delta(\boldsymbol{X})$ and $\boldsymbol{\tau}^\delta(\boldsymbol{Y}) \in \boldsymbol{\tau}^\delta(\boldsymbol{Y})$, one has

$$
\Lambda(\|\boldsymbol{\tau}^\delta(\boldsymbol{Y}) - \boldsymbol{\tau}^\delta(\boldsymbol{X})\|) \leq \sum_{i=1}^n \tau_i^{\boldsymbol{Y}} - \sum_{i=1}^n \tau_i^{\boldsymbol{X}} = \rho_l^\delta(\boldsymbol{Y}) - \rho_l^\delta(\boldsymbol{X}) \leq \frac{L}{\delta}\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1].
$$

Consequently

$$
\|\boldsymbol{\tau}^\delta(\boldsymbol{Y}) - \boldsymbol{\tau}^\delta(\boldsymbol{X})\| \leq \Lambda^{-1}\left(\frac{L}{\delta}\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1]\right).
$$

Thus,

$$
\|\boldsymbol{\tau}^\delta(\boldsymbol{Y}) - \boldsymbol{\tau}^\delta(\boldsymbol{X})\| \leq \Lambda^{-1}\left(\frac{L}{\delta}\mathbb{E}[\|\boldsymbol{X} - \boldsymbol{Y}\|_1]\right).
$$

By swapping the positions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, we obtain (6.4). This completes the proof. ∎

B.10 Proof of Theorem 6.3

Proof. Following the proof of Theorem 4.1, we can recast the MRLM problem (2.1) as

$$\rho_l^\delta(P) = \inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\sup_{\lambda\ge 0}\left\{(1-\lambda\delta)\sum_{i=1}^n \tau_i + \lambda\int_{\mathbb{R}^n} l(\boldsymbol{x}-\boldsymbol{\tau})P(d\boldsymbol{x})\right\}. \tag{B.6}$$

Since $P \in \mathcal{M}_n^l$, then $\int_{\mathbb{R}^n} l(\boldsymbol{x}-\boldsymbol{\tau})P(d\boldsymbol{x})$ is finite-valued. By Theorem 4.1, $\rho_l^\delta(P) \in \mathbb{R}$. Moreover, since the Slater condition is satisfied, then the strong duality holds and the Lagrange multiplier λ is bounded, see, e.g., Pomerol (1981). Furthermore, from the proof of Theorem 4.1, we have $1-\lambda\delta > 0$, i.e., $\lambda < \frac{1}{\delta}$. Consequently, we can write (B.6) as

$$\rho_l^\delta(P) = \sup_{\lambda\in[0,\frac{1}{\delta}]}\inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\left\{(1-\lambda\delta)\sum_{i=1}^n \tau_i + \lambda\int_{\mathbb{R}^n} l(\boldsymbol{x}-\boldsymbol{\tau})P(d\boldsymbol{x})\right\}. \tag{B.7}$$

Likewise, we have

$$\rho_l^\delta(P^N) \quad = \quad \sup_{\lambda\in[0,\frac{1}{\delta}]}\inf_{\boldsymbol{\tau}\in\mathbb{R}^n}\left\{(1-\lambda\delta)\sum_{i=1}^n \tau_i + \lambda\frac{1}{N}\sum_{i=1}^N l(\boldsymbol{x}_i-\boldsymbol{\tau})\right\}. \tag{B.8}$$

Since l is Lipschitz continuous with modulus L, then

$$|\rho_l^\delta(P^N)-\rho_l^\delta(Q^N)| \le \frac{1}{\delta N}\sup_{\boldsymbol{\tau}\in\mathbb{R}^n}\left|\sum_{i=1}^N l(\boldsymbol{x}_i-\boldsymbol{\tau}) - \sum_{i=1}^N l(\widetilde{\boldsymbol{x}}_i-\boldsymbol{\tau})\right| \le \frac{1}{N}\cdot\frac{L}{\delta}\sum_{i=1}^n |\boldsymbol{x}_i-\widetilde{\boldsymbol{x}}_i|. \tag{B.9}$$

By virtue of Wang et al. (2021, Theorem 4.5) or Guo and Xu (2021, Theorem 3), we have

$$\mathsf{dl}_K(P^{\otimes N}\circ\widehat{\varrho}_N^{-1}, Q^{\otimes N}\circ\widehat{\varrho}_N^{-1}) \le \frac{L}{\delta}\mathsf{dl}_K(P,Q). \tag{B.10}$$

This completes the proof. ■