Monotonic mean-deviation risk measures

Xia Han[*] Ruodu Wang[†] Qinyu Wu[‡]

20th February 2025

Abstract

We propose and study the class of monotonic mean-deviation risk measures, represented by a combination of a risk-weighted deviation functional and the expectation. These risk measures belong to the class of consistent risk measures and admit an axiomatic characterization via preference relations. By further assuming the convexity and linearity of the risk-weighting function, we obtain convex and coherent risk measures among this class, giving rise to many new explicit examples of convex and nonconvex consistent risk measures. In particular, we specialize in the convex case of the monotonic mean-deviation measure and obtain its dual representation. Further, we establish asymptotic consistency and normality of the natural estimators of the monotonic mean-deviation measures. Finally, monotonic mean-deviation measures are applied to the problem of portfolio selection using financial data.

KEYWORDS: Risk management, axiomatization, deviation measures, monotonicity, convexity

1 Introduction

In the last few decades, risk measures and deviation measures have been popular in banking and finance for various purposes, such as calculating solvency capital reserves, pricing of insurance risks, performance analysis, and internal risk management. Roughly speaking, deviation measures evaluate the degree of nonconstancy in a random variable (i.e., the extent to which outcomes may deviate from a center, such as the expectation of the random variable), whereas risk measures evaluate overall prospective loss (from the benchmark of zero loss). Different classes of axioms are proposed for risk measures and deviation measures in the literature; see Artzner et al. (1999) for coherent risk measures, Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002) for convex risk measures, and Rockafellar et al. (2006) for generalized deviation measures.

Since the seminal work of Markowitz (1952), mean-deviation or mean-risk problems have been central to financial studies. In this context, a decision maker's objective functional U on a loss/profit random variable X can be characterized by

$$U(X) = V(\mathbb{E}[X], D(X)), \tag{1}$$

where $\mathbb{E}$ is the expectation, V is a monotonic bivariate function, and D measures the *risk part* of X, which is chosen as the variance in the context of Markowitz (1952), and as a risk measure or deviation

[*]School of Mathematical Sciences and LPMC, Nankai University, Tianjin, China. E-mail: xiahan@nankai.edu.cn

[†]Department of Statistics and Actuarial Science, University of Waterloo, Canada. E-mail: wang@uwaterloo.ca

[‡]Correspondence: Department of Statistics and Actuarial Science, University of Waterloo, Canada. E-mail: q35wu@uwaterloo.ca

measure in subsequent studies. For instance, the classic problem of expected return maximization with variance constraint can be written as to minimize $V_\sigma(\mathbb{E}[X], \mathrm{Var}(X))$ where

$$V_\sigma(m,d) := m + \infty \times \mathbb{1}_{\{d>\sigma^2\}} \tag{2}$$

for some $\sigma > 0$,[1] and it is typically solved by minimizing $V^\lambda(\mathbb{E}[X], \mathrm{Var}(X))$, where

$$V^\lambda(m,d) := m + \lambda d \tag{3}$$

for some $\lambda > 0$ via a Lagrangian method. Because any law-invariant coherent risk measure R induces a deviation measure D via $D = R - \mathbb{E}$, we can write the mean-risk problem with a coherent risk measure R as

$$V(\mathbb{E}[X], R(X)) = V'(\mathbb{E}[X], D(X)),$$

where $V'(m,d) = V(m, d+m)$. Therefore, in this paper we focus on (1) with D being a deviation measure.

The mean-deviation model is widely used in the finance and optimization literature; for example, the early work of Markowitz (1952), Sharpe (1964), and Simaan (1997), and the more recent progress in Grechuk et al. (2012), Grechuk and Zabarankin (2012), Rockafellar and Uryasev (2013), and Herdegen and Khan (2022a, 2024). Nevertheless, only few studies, including Grechuk et al. (2012), have focused on the preference functional U in (1), which is an interesting mathematical object by itself, as the decision criterion used for optimization.

In general, U in (1) is not monotonic, as mean-variance analysis is inconsistent with monotonic preferences; see, e.g., Maccheroni et al. (2009). Monotonicity is self-explanatory and is common in the literature on decision theory and risk measures. As of today, the most popular risk measures are *monetary risk measures* that satisfy the two properties of monotonicity and cash additivity, with *Value-at-Risk* (VaR) and *Expected Shortfall* (ES) being the most famous examples. The monetary property allows for the interpretation of a risk measure as a regulatory capital requirement defined via acceptance sets. Therefore, it is natural to consider the intersection of mean-deviation models and monetary risk measures, enjoying the advantages of both streams of literature. The functionals belonging to both classes will be called *monotonic mean-deviation (risk) measures*. We omit the term "risk" for simplicity, while keeping in mind that these functionals are risk measures in the sense of Artzner et al. (1999) and Föllmer and Schied (2016).

Throughout, we consider deviation measures D satisfying the properties of Rockafellar et al. (2006) (defined in Section 2). A natural candidate for monotonic mean-deviation measures is to use the sum $U = \mathbb{E} + \lambda D$ for some $\lambda \geqslant 0$, which appears in the Markowitz model through (3) and also in insurance pricing (see Denneberg (1990) and Furman and Landsman (2006)). However, this is not the only possible choice. In Section 3, we characterize monotonic mean-deviation measures among general mean-deviation models (Theorem 1). It turns out that they admit the form of a combination of the expectation and a deviation part distorted by a *risk-weighting function* g, and D needs to satisfy a condition of range normalization (defined in Section 2). Such monotonic mean-deviation measures are denoted by MD_g^D, that is,

$$\mathrm{MD}_g^D = g \circ D + \mathbb{E}. \tag{4}$$

To the best of our knowledge, the form of risk measures in (4) has not been proposed in the literature, except for some special cases. Although measuring both the mean and the diversification (via the deviation measure), MD_g^D is not necessarily a convex risk measure in the sense of Föllmer and Schied (2016). Nevertheless, MD_g^D satisfies a weaker requirement reflecting on diversification, that

[1]Here we interpret X as a loss, so the expected return is $-\mathbb{E}[X]$.

is, consistency with respect to second-order stochastic dominance. Compared with $U = \mathbb{E} + \lambda D$, the risk-weighting function g allows us to relax restrictions of the mean-deviation model, in a way similar to Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002), which relaxed coherent risk measures to convex ones, and to Castagnoli et al. (2022), who relaxed convex risk measures to star-shaped ones. Thus, the new class of risk measures offers additional flexibility while maintaining the essential ingredients needed to assess risk via deviation in particular contexts.

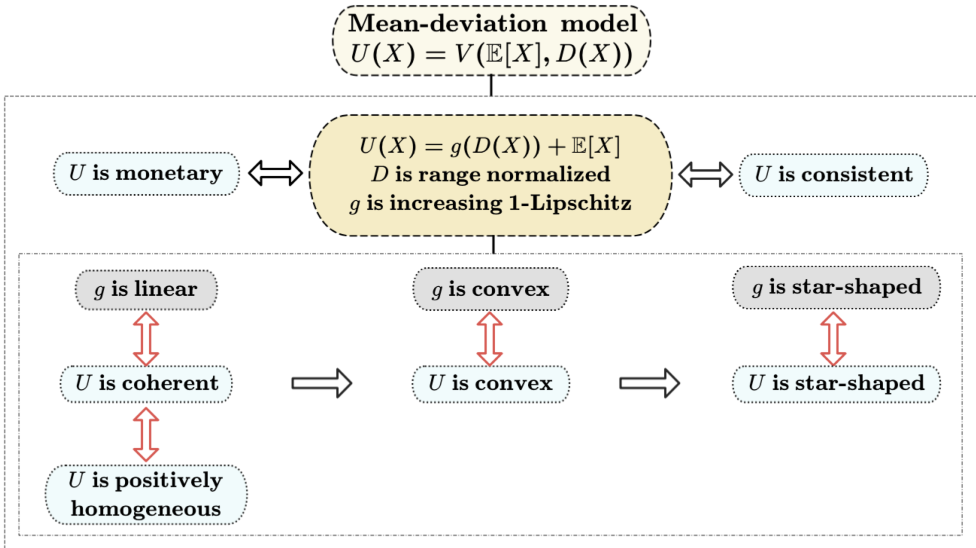


Figure 1: An illustration of the properties of the mean-deviation model

In addition to proposing the mean-deviation measures in (4), our main contributions include a comprehensive study on this class of risk measures. In Section 4, an axiomatic foundation for MD_g^D (Theorem 2) is proposed on the basis of the results of Grechuk et al. (2012), and characterizations for coherent, convex or star-shaped risk measures are obtained in Theorem 3. We show that there is a one-to-one correspondence between the MD_g^D and the risk-weighting function g, and hence the above classes can be identified based on properties of g. Figure 1 contains an illustration of the properties of MD_g^D. In particular, the convexity of g is equivalent to the convexity of MD_g^D. As a consequence, our results offer new convex risk measures with explicit formulas, in addition to the existing convex distortion risk measures and entropy risk measures; see e.g., Dhaene et al. (2006), Laeven and Stadje (2013) and Föllmer and Schied (2016). Specifically, these formulas help us construct risk measures that are consistent yet not convex, or convex but not coherent (Theorem 3 and Proposition 2).

In Section 5, we specialize in convex monotonic mean-deviation measure and further study the dual representation of MD_g^D (Theorem 4), which is obtained directly through the conjugate function of g. In Section 6, when the deviation measures are the convex signed Choquet integral defined in Wang et al. (2020b), we discuss the non-parametric estimation of MD_g^D (Theorem 5). The asymptotic normality and the asymptotic variance for the empirical estimators are obtained explicitly. These results yield an intuitive trade-off between statistical efficiency, in terms of estimation error, and sensitivity to risk, in terms of the risk-weighting function. In Section 7, we present an application of MD_g^D in portfolio selection based on financial data, and discuss some empirical observations.

We conclude the paper in Section 8. Supplementary materials containing the characterization of monotonicity in mean-deviation models are put in Appendix A, and the details in the axiomatization results are relegated to Appendix B. Appendix C provides a proof that is omitted from Section 6. Appendix D further analyzes worst-case values of MD_g^D under two popular settings to show its feasibility in model uncertainty problems.

2 Preliminaries

Throughout this paper, we work with a nonatomic probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{X}$ be a convex cone of random variables containing all constants. All the equalities and inequalities of functionals on $(\Omega, \mathcal{F}, \mathbb{P})$ are under $\mathbb{P}$ almost surely ($\mathbb{P}$-a.s.) sense. Let $X \in \mathcal{X}$ represent the random loss faced by financial institutions in a fixed period of time. That is, a positive value of $X \in \mathcal{X}$ represents a loss and a negative value represents a surplus in our sign convention, which is used by, e.g., McNeil et al. (2015). Further, denote by $\mathcal{X}^\circ$ the set of all nonconstant random variables in $\mathcal{X}$. Let F_X be the distribution function of X, and we write $X \stackrel{\mathrm{d}}{=} Y$ if two random variables X and Y have the same distribution. Terms such as increasing or decreasing functions are in the non-strict sense. For $p \in [1, \infty)$, we denote by $L^p = L^p(\Omega, \mathcal{F}, \mathbb{P})$ the set of all random variables X such that $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p} < \infty$. Furthermore, $L^\infty = L^\infty(\Omega, \mathcal{F}, \mathbb{P})$ is the space of all essentially bounded random variables, and $L^0 = L^0(\Omega, \mathcal{F}, \mathbb{P})$ denotes the space of all random variables. When considering a mapping defined on L^p for some $p \in [1, \infty]$, we refer to its continuity with respect to the L^p-norm. For a real function g, we use g' to denote the left derivative of g, whenever it exists. For a convex or concave function, the left derivative always exists on the inner of its domain (see e.g., Proposition A.4 of Föllmer and Schied (2016)). Denote by $x_+ = \max\{0, x\}$ for $x \in \mathbb{R}$, and let $\mathbb{1}_A$ be the indicator function of A.

We define two important risk measures in banking and insurance practice. The Value-at-Risk (VaR) at level $\alpha \in (0, 1)$ is the functional $\mathrm{VaR}_\alpha : L^0 \to \mathbb{R}$ defined by

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

which is precisely the left α-quantile of X. In some contexts, we also use $F_X^{-1}(\alpha)$ instead of $\mathrm{VaR}_\alpha(X)$ for convenience. The Expected Shortfall (ES) at level $\alpha \in [0, 1)$ is the functional $\mathrm{ES}_\alpha : L^1 \to \mathbb{R}$ defined by

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

Artzner et al. (1999) introduced *coherent risk measures* as functionals $\rho : \mathcal{X} \to (-\infty, \infty]$ that satisfy the following four properties.

[M] Monotonicity: $\rho(X) \leqslant \rho(Y)$ for all $X, Y \in \mathcal{X}$ with $X \leqslant Y$.

[CA] Cash additivity: $\rho(X + c) = \rho(X) + c$ for all $c \in \mathbb{R}$ and $X \in \mathcal{X}$.

[PH] Positive homogeneity: $\rho(\lambda X) = \lambda \rho(X)$ for all $\lambda \in (0, \infty)$ and $X \in \mathcal{X}$.

[SA] Subadditivity: $\rho(X + Y) \leqslant \rho(X) + \rho(Y)$ for all $X, Y \in \mathcal{X}$.

ES satisfies all four properties above, whereas VaR does not satisfy [SA]. We say that a functional ρ is *monetary* if it satisfies [M] and [CA]. Moreover, ρ is a *convex risk measure* if it is monetary and further satisfies

[Cx] Convexity: $\rho(\lambda X + (1-\lambda)Y) \leqslant \lambda \rho(X) + (1-\lambda)\rho(Y)$ for all $X, Y \in \mathcal{X}$ and $\lambda \in [0, 1]$.

Clearly, [PH] together with [SA] implies [Cx]. Risk measures that satisfy [CA] and [Cx] but not [M] have been studied by e.g., Filipović and Svindland (2008). For more discussions and interpretations of these properties, we refer to Föllmer and Schied (2002) and Frittelli and Rosazza Gianin (2002). Another class of risk measures is defined based on consistency with respect to second-order stochastic dominance (SSD):

[SC] SSD-consistency: $\rho(X) \leqslant \rho(Y)$ if $X \leqslant_{\mathrm{SSD}} Y$ (i.e., $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all increasing convex functions u).[2]

The monetary risk measures that satisfy [SC] are called *consistent risk measures*, and were introduced by Dana (2005). Consistent risk measures are characterized by Mao and Wang (2020). The property [SC] is often called *strong risk aversion* for a preference functional in decision theory; see Rothschild and Stiglitz (1970). A related notion to SSD is convex order, denoted by $\leqslant_{\mathrm{cx}}$, which is also called mean-preserving spread, and for $X, Y \in \mathcal{X}$, $X \leqslant_{\mathrm{cx}} Y$ means $X \leqslant_{\mathrm{SSD}} Y$ and $\mathbb{E}[X] = \mathbb{E}[Y]$.

In decision making, deviation measures are also introduced to measure variability associated with a random variable, and are systematically studied for their applications to risk management in areas such as portfolio optimization and engineering. Such measures include standard deviation as a special case but need not be symmetric with respect to gains and losses. Deviation measures of Rockafellar et al. (2006) are formally defined below.

Definition 1 (Deviation measures)**.** Fix $p \in [1, \infty]$. A *deviation measure* is a functional $D : L^p \to [0, \infty)$ satisfying

(D1) $D(X + c) = D(X)$ for all $X \in L^p$ and $c \in \mathbb{R}$.

(D2) $D(X) > 0$ for all $X \in (L^p)^\circ$.

(D3) $D(\lambda X) = \lambda D(X)$ for all $X \in L^p$ and $\lambda \geqslant 0$.

(D4) $D(X + Y) \leqslant D(X) + D(Y)$ for all $X, Y \in L^p$.

From the definition, it is straightforward to see that a deviation measure D is convex and satisfies $D(c) = 0$ for all $c \in \mathbb{R}$. Rockafellar et al. (2006) defined deviation measures on L^2 because it gives access to tools associated with duality. However, as mentioned in Rockafellar et al. (2006), this does not prevent us from working with the general L^p norms for $p \in [1, \infty]$. Moreover, we will focus on *law-invariant deviation measures*, which further satisfy

(D5) $D(X) = D(Y)$ for all $X, Y \in L^p$ if $X \stackrel{d}{=} Y$.

Law-invariant deviation measures include, for instance, standard deviation, semideviation, ES deviation and range-based deviation; see Examples 1 and 2 of Rockafellar et al. (2006) and Section 4.1 of Grechuk et al. (2012). We use $\mathcal{D}^p$ to denote the set of all law-invariant deviation measures.

A deviation measure D is *upper range-dominated* if it has the following property

$$D(X) \leqslant \text{ess-sup} X - \mathbb{E}[X] \quad \text{for all } X \in L^p, \tag{5}$$

where ess-sup$X$ is the essential supremum of X. For more discussions and interpretations of the properties of deviation measures mentioned above, we refer to Rockafellar et al. (2006).

Next, we introduce a new property that will be used throughout the rest of the paper. We say a deviation measure $D \in \mathcal{D}^p$ is *ranged-normalized* if

$$\sup_{X \in (L^p)^\circ} \frac{D(X)}{\text{ess-sup} X - \mathbb{E}[X]} = 1.$$

[2] Note that a random variable represents the random loss instead of the random wealth. In our context, SSD is also known as increasing convex order in probability theory and stop-loss order in actuarial science. Up to a sign change that converts losses to gains, SSD corresponds to increasing concave order, which is the classic second-order stochastic dominance in decision theory.

The set of all range-normalized deviation measures on L^p is denoted by $\overline{\mathcal{D}}^p$, which has the form:

$$\overline{\mathcal{D}}^p = \left\{ D \in \mathcal{D}^p : \sup_{X \in (L^p)^\circ} \frac{D(X)}{\text{ess-sup} X - \mathbb{E}[X]} = 1 \right\}. \tag{6}$$

For $D \in \mathcal{D}^p$, the condition $\lambda D \in \overline{\mathcal{D}}^p$ for some $\lambda > 0$ is called *weak upper range dominance* by Grechuk et al. (2012). It is clear from (5) that every upper range-dominated deviation measure is weakly upper range-dominated. In particular, if D takes the form of $\mathrm{ES}_\alpha - \mathbb{E}$ with $\alpha \in (0, 1)$, ess-sup$X - \mathbb{E}[X]$ or $\mathbb{E}[|X - \mathbb{E}[X]|]/2$, we have $D \in \overline{\mathcal{D}}^p$ (see Example 5 of Grechuk et al. (2012) for the last one). The class of weakly upper range-dominated deviation measures also includes the mean-absolute deviation, the Gini deviation, the inter-ES range, and the inter-expectile range (for the last two, see Bellini et al. (2022)).

The deviation measures are closely connected to coherent risk measures. It is shown in Theorem 2 of Rockafellar et al. (2006) that upper range bounded deviation measures D correspond one-to-one with coherent, strictly expectation bounded[3] risk measures R with the relation that $R = D + \mathbb{E}$. Additionally, note that $R = D + \mathbb{E}$ is a finite coherent risk measure on L^p for any $D \in \overline{\mathcal{D}}^p$. It follows that R is continuous (see e.g., Corollary 2.3 of Kaina and Rüschendorf (2009)) and so is D. Below we provide a characterization of range-normalized deviation measures based on coherent risk measures.

Proposition 1. *Fix $p \in [1, \infty]$. The deviation measure $D \in \mathcal{D}^p$ is range-normalized if and only if $D + \mathbb{E}$ is a coherent risk measure and $\lambda D + \mathbb{E}$ is not a coherent risk measure for $\lambda > 1$.*

Proof. The necessity follows immediately from Theorem 2 of Rockafellar et al. (2006) since $D \in \overline{\mathcal{D}}^p$ is upper range-dominated and λD is not upper range-dominated for any $\lambda > 1$. Conversely, we assume by contradiction that D is not range-normalized. Then, either $kD \in \overline{\mathcal{D}}^p$ for some $k > 1$ or $kD \in \overline{\mathcal{D}}^p$ for some $k < 1$ holds.

In the first case, there exists $\lambda > 1$ such that λD is upper range-dominated. Applying Theorem 2 of Rockafellar et al. (2006), we have that $\lambda D + \mathbb{E}$ is a coherent risk measure, thereby leading to a contradiction. In the second case, it holds that $D + \mathbb{E}$ is not a coherent risk measure since D is not upper range-dominated, which also yields a contradiction. □

The additive structure $\lambda D + \mathbb{E}$ can be seen as a special form of the combination of mean and deviation. Below, we define a class of general mean-deviation models that is not necessarily an additive form.

Definition 2 (Mean-deviation model)**.** Fix $p \in [1, \infty]$. For a deviation measure $D \in \mathcal{D}^p$, a *mean-deviation model* is a functional $U : L^p \to (-\infty, \infty]$ defined as

$$U(X) = V(\mathbb{E}[X], D(X)), \tag{7}$$

where $V : \mathbb{R} \times [0, \infty) \to (-\infty, \infty]$ satisfies (i) V is increasing component-wise; (ii) $V(m, 0) = m$ for all $m \in \mathbb{R}$; (iii) $V(m, d)$ is not determined only by m.[4]

The three conditions on V in Definition 2 are simple and intuitive. More specifically, (i) is the basic requirement that U increases when the mean or deviation increases, with the other argument fixed; (ii) means that a constant random variable has risk value equal to itself; and (iii) means that the model is not trivial in the sense that it does not ignore the deviation $D(X)$. Our definition is different from that of Grechuk et al. (2012), who required further strict monotonicity of V with a real-valued

[3] A risk measure $\rho : \mathcal{X} \to (-\infty, \infty]$ is *strictly expectation bounded* if it satisfies $\rho(X) > \mathbb{E}[X]$ for all $X \in \mathcal{X}^\circ$.

[4] That is, there exist $m \in \mathbb{R}$ and $d_1, d_2 \geqslant 0$ such that $V(m, d_1) \neq V(m, d_2)$.

range. Therefore, our requirement is weaker than that of Grechuk et al. (2012), and this relaxation allows us to include the most popular models of Markowitz (1952) in (2), that is,

$$V(\mathbb{E}[X], \mathrm{SD}(X)) = V_\sigma(\mathbb{E}[X], \mathrm{Var}(X)) = \mathbb{E}[X] + \infty \times \mathbb{1}_{\{\mathrm{SD}(X)>\sigma\}},$$

which is neither strictly increasing nor real-valued. Here we use SD (the standard deviation) instead of Var because $\mathrm{SD} \in \mathcal{D}^2$.

As mentioned in the Introduction, the mean-deviation model has many good properties; however, it is not necessarily monotonic or cash additive in general, and thus is not a monetary risk measure. Grechuk et al. (2012) provided an axiomatic framework for the mean-deviation model via the preference relation by further assuming some other properties; but this framework does not belong to the class of monetary risk measures. Han et al. (2023) characterized mean-deviation models with $D = \mathrm{ES}_\alpha - \mathbb{E}$ for $\alpha \in (0, 1)$ by extending axioms for ES in Wang and Zitikis (2021), which can be further required to be monetary. Since [M] and [CA] are common in the literature concerning decision theory and risk measures and correspond to the interpretation of a risk measure as a regulatory capital requirement, it is natural to further consider general conditions for a mean-deviation model to be monetary. This leads to the main object of this paper, monotonic mean-deviation measures, which are formally introduced in the next section.

3 Monotonic mean-deviation measures

In this section, we derive an explicit representation of mean-deviation models that are monetary. To this end, we revisit the set of range-normalized deviation measures defined in (6) and introduce a continuity condition for a real function g, i.e., g is *λ-Lipschitz continuous* for some $\lambda > 0$ if

$$|g(x) - g(y)| \leqslant \lambda |x - y| \quad \text{for } x, y \text{ in the domain of } g. \tag{8}$$

The following lemma shows that a necessary condition for property [M] of a mean-deviation model is that the deviation measure satisfies weak upper range dominance. It is useful in the proof of Theorem 1, which is the main result in this section.

Lemma 1. *Fix $p \in [1, \infty]$, and let $D \in \mathcal{D}^p$. If $U = V(\mathbb{E}, D)$ in (7) satisfies [M], then we have $U(X) < \infty$ for all $X \in L^p$, and there exists $\lambda > 0$ such that $\lambda D \in \overline{\mathcal{D}}^p$.*

Proof. To show that $U(X)$ is finite for $X \in L^p$, take $Y \in L^\infty$ such that $\mathbb{E}[Y] = \mathbb{E}[X]$ and $D(Y) = D(X)$. Such Y exists because D is positively homogeneous. Therefore, $U(X) = U(Y) \leqslant U(\text{ess-sup} Y) = V(\text{ess-sup} Y, 0) = \text{ess-sup} Y < \infty$.

We next prove that

$$K := \sup_{X \in (L^p)^\circ} \frac{D(X)}{\text{ess-sup} X - \mathbb{E}[X]} < \infty. \tag{9}$$

by contradiction, which is equivalent to $\lambda D \in \overline{\mathcal{D}}^p$ with $\lambda = 1/K$. Assume that $K = \infty$ in (9). For $X_1, X_2 \in L^p$ such that $\mathbb{E}[X_1] < \mathbb{E}[X_2]$, let $m_1 = \mathbb{E}[X_1]$, $d_1 = D(X_1)$, $m_2 = \mathbb{E}[X_2]$, $d_2 = D(X_2)$, and $e = m_2 - m_1$. If $K = \infty$, there exists Y_1 such that $D(Y_1)/(\text{ess-sup} Y_1 - \mathbb{E}[Y_1]) \geqslant d_1/e$. Denote by $Y_2 = e(Y_1 - \text{ess-sup} Y_1)/(\text{ess-sup} Y_1 - \mathbb{E}[Y_1]) + m_2$. It holds that $\mathbb{E}[Y_2] = -e + m_2 = m_1$ and $D(Y_2) \geqslant d_1$, and thus $U(X_1) = V(m_1, d_1) \leqslant V(\mathbb{E}[Y_2], D(Y_2)) = U(Y_2)$. On the other hand, observe that $Y_2 \leqslant m_2$. Consequently, by monotonicity, we have $U(Y_2) \leqslant U(m_2)$. Thus, we have $U(X_1) \leqslant U(Y_2) \leqslant U(m_2) \leqslant U(X_2)$, which implies that $U(X) \leqslant U(Y)$ for every X and Y with $\mathbb{E}[X] < \mathbb{E}[Y]$. Hence,

$$\mathbb{E}[X] - \varepsilon = U(\mathbb{E}[X] - \varepsilon) \leqslant U(X) \leqslant U(\mathbb{E}[X] + \varepsilon) = \mathbb{E}[X] + \varepsilon$$

for any $X \in L^p$ and $\varepsilon > 0$. Letting $\varepsilon \downarrow 0$ yields $U(X) = \mathbb{E}[X]$, contradicting (iii) in Definition 2. Therefore, we conclude that $K < \infty$. This completes the proof. □

Next, we establish a representation result for mean-deviation models that are monetary below.

Theorem 1. *Fix $p \in [1, \infty]$. Suppose that $U : L^p \to (-\infty, \infty]$ is a mean-deviation model in (7) with $D \in \mathcal{D}^p$. The following statements are equivalent.*

(i) U is a monetary risk measure.

(ii) U is a consistent risk measure.

(iii) For some $\lambda > 0$, $\lambda D \in \overline{\mathcal{D}}^p$ and $U = g \circ D + \mathbb{E}$ where $g : [0, \infty) \to \mathbb{R}$ is a non-constant increasing and λ-Lipschitz continuous function satisfying $g(0) = 0$.

Proof. (ii) ⇒ (i) is trivial.

(iii) ⇒ (ii): Without loss of generality we can take $\lambda = 1$. The property of [CA] is clear. Next, we verify the property of [M]. For any $X, Y \in L^p$ with $X \leqslant Y$, we have $U(X) \leqslant U(Y)$ if $D(Y) \geqslant D(X)$. Suppose now that $D(X) > D(Y)$. It holds that

$$U(Y) - U(X) = g(D(Y)) + \mathbb{E}[Y] - g(D(X)) - \mathbb{E}[X] \geqslant D(Y) - D(X) + \mathbb{E}[Y] - \mathbb{E}[X],$$

where use the 1-Lipschitz continuity of g in the inequality. Since $D \in \overline{\mathcal{D}}^p$, it follows from Theorem 2 of Rockafellar et al. (2006) that there exists one-to-one correspondence with coherent risk measures denoted by R in the relation that $R(X) = D(X) + \mathbb{E}[X]$ for $X \in L^p$. The monotonicity of R implies that

$$U(Y) - U(X) \geqslant D(Y) - D(X) + \mathbb{E}[Y] - \mathbb{E}[X] = R(Y) - R(X) \geqslant 0.$$

Hence, we have verified [M] of U.

It remains to be shown that U satisfies [SC]. We recall that $D \in \overline{\mathcal{D}}_p$ is continuous because $R = D + \mathbb{E}$ is a finite coherent risk measure on L^p, which is continuous (see e.g., Corollary 2.3 of Kaina and Rüschendorf (2009)). Since D also satisfies convexity and law-invariance, and the space is nonatomic, we obtain that D is consistent with respect to the convex order (see, e.g., Theorem 4.1 of Dana (2005)), and the same property holds for U because of the increasing monotonicity of g. Combining with [M] of U, it follows from Theorem 4.A.6 of Shaked and Shanthikumar (2007) that U satisfies [SC]. This completes the proof of (iii) ⇒ (ii).

(i) ⇒ (iii): Define $g(d) = V(0, d)$ for $d \geqslant 0$. It is clear that g is an increasing function with $g(0) = V(0, 0) = 0$. By [CA], we have

$$\begin{aligned} U(X) &= U(X - \mathbb{E}[X]) + \mathbb{E}[X] \\ &= V(0, D(X)) + \mathbb{E}[X] = g(D(X)) + \mathbb{E}[X]. \end{aligned}$$

By Lemma 1 below, we have $\lambda D \in \overline{\mathcal{D}}^p$ for some $\lambda > 0$. It remains to be shown that g is λ-Lipschitz continuous. Denote by $k = 1/\lambda$. Since $\lambda D \in \overline{\mathcal{D}}^p$, for any $\varepsilon \in (0, k)$, there exists X_1 such that

$$k - \varepsilon < \frac{D(X_1)}{\text{ess-sup}X_1 - \mathbb{E}[X_1]} \leqslant k.$$

For any $a > 0$ and $d \geqslant 0$, define

$$X_2 = a\frac{X_1 - \text{ess-sup}X_1}{\text{ess-sup}X_1 - \mathbb{E}[X_1]} \quad \text{and} \quad X_3 = \frac{d}{a}X_2 + d.$$

It is obvious that $\mathbb{E}[X_2] = -a, X_2 \leqslant 0$ and $\mathbb{E}[X_3] = 0$. Moreover, $a(k-\varepsilon) < D(X_2) \leqslant ak$ and $d(k-\varepsilon) < D(X_3) \leqslant dk$. Additionally, $\mathbb{E}[X_2 + X_3] = -a$ and $(d+a)(k-\varepsilon) < D(X_2+X_3) \leqslant (d+a)k$. Since $X_2 + X_3 \leqslant X_3$, by [M], we have $g(D(X_2+X_3)) + \mathbb{E}[X_2+X_3] \leqslant g(D(X_3)) + \mathbb{E}[X_3]$. Letting $\varepsilon \to 0$, we conclude that $g_-((d+a)k) \leqslant g(dk) + a$, where $g_-(x) = \lim_{y\uparrow x} g(y)$ for all $x \geqslant 0$. This is equivalent to $g_-(d+a) - g(d) \leqslant \lambda a$ for any $a > 0$ and $d \geqslant 0$. Note that g is increasing. We have that $g : [0,\infty) \to \mathbb{R}$ is λ-Lipschitz continuous. This completes the proof. □

Theorem 1 leads to the following definition of monotonic mean-deviation measures.

Definition 3. Fix $p \in [1,\infty]$ and let $D \in \overline{\mathcal{D}}^p$. A *monotonic mean-deviation measure* $\mathrm{MD}_g^D : L^p \to \mathbb{R}$ is defined by

$$\mathrm{MD}_g^D(X) = g(D(X)) + \mathbb{E}[X], \tag{10}$$

where $g : [0,\infty) \to \mathbb{R}$ is a non-constant increasing and 1-Lipschitz continuous function satisfying $g(0) = 0$, called a *risk-weighting function*. We use $\mathcal{G}$ to denote the set of such functions g.

The functionals defined in Definition 3 precisely encompass all mean-deviation models that are monetary. Specifically, if D and g satisfy the conditions in Theorem 1 (iii), then there exists $\widetilde{D} \in \overline{\mathcal{D}}^p$ and $\widetilde{g} \in \mathcal{G}$ such that $\widetilde{g} \circ \widetilde{D} = g \circ D$. This means that for any pair (D, g) satisfying the conditions in Theorem 1 (iii), there is a corresponding pair $(\widetilde{D}, \widetilde{g})$ that satisfies the conditions in Definition 3. Therefore, monotonic mean-deviation measures and precisely mean-deviation models that are monetary.

The interpretation of g should be self-evident: it dictates how $D(X)$ is reflected in the calculation of MD_g^D, and it is a generalization of the risk-weighting parameter λ in $\mathbb{E} + \lambda D$, hence the name. The reason for requiring the conditions $D \in \overline{\mathcal{D}}^p$ and $g \in \mathcal{G}$ in Definition 3 has been justified in Theorem 1. Intuitively, since $D \in \overline{\mathcal{D}}^p$ is not monotonic, while $\mathbb{E}$ is monotonic, a function g that satisfies the 1-Lipschitz continuity can regularize the influence of D on monotonicity, ensuring the overall mean-deviation model is monotonic. Note that MD_g^D is not subadditive in general. For instance, let $g(x) = (x-1)_+$ and $D = \mathrm{ES}_\alpha - \mathbb{E}$ with $\alpha = 0.5$. Take X, Y be such that $\mathbb{P}(X = Y = 0) = \mathbb{P}(X = Y = 2) = 1/2$. One can check $D(X) = D(Y) = 1$ and $D(X+Y) = 2$. Thus, $\mathrm{MD}_g^D(X+Y) = 3$ and $\mathrm{MD}_g^D(X) = \mathrm{MD}_g^D(Y) = 1$, which violates subadditivity. In practice, a simple class of functions g can be chosen as $g(x) = \lambda(x-\theta)_+$ for some $\theta \geqslant 0$ and $\lambda \in (0,1]$. The interpretation is clear: Risks X with deviation $D(X)$ smaller than θ are seen as not very dangerous and are assessed by their expected value. Risks X with deviation $D(X)$ larger than θ are dangerous and penalized in their risk assessment. In portfolio management, the parameters θ and λ can be calibrated on the basis of the performance of the risk measure on test data.

The expected return maximization with variance constraint of Markowitz (1952) has the form MD_g^D where $D = \mathrm{SD}$ and $g(d) = \infty \times \mathbb{1}_{\{d>\sigma\}}$ for some $\sigma > 0$ as in (2). In this example, g is not real-valued. Therefore, although sharing the form (10), MD_g^D is not a monotonic mean-deviation measure. Similarly, for $\lambda > 0$, the functional $\mathrm{MD}_g^D(X) = \lambda(\mathrm{SD}(X))^2 + \mathbb{E}[X]$ in (3) or $\mathrm{MD}_g^D(X) = \lambda \mathrm{SD}(X) + \mathbb{E}[X]$ is not a monotonic mean-deviation measure, because SD does not satisfies (9) for any $p \in [1,\infty]$. Nevertheless, in all three examples, g is convex. Indeed, convexity of g has important implications, and this will be studied in Section 4.2 below.

4 Characterization

4.1 Axiomatization of monotonic mean-deviation measures

In this section, we present an axiomatization of the monotonic mean-deviation measure MD_g^D through preference relations. This axiomatization is very similar to that of Grechuk et al. (2012), who axiomatized preferences represented by a mean-deviation model $X \mapsto V(\mathbb{E}[X], D(X))$ satisfying [M] with some strictly increasing function V. We relegate all the details, including all the proofs and a comparison with Grechuk et al. (2012), to Appendix B. Our main purpose here is to show that MD_g^D has an axiomatic foundation.

A preference relation $\succeq$ is defined as a total preorder.[5] As usual, $\succ$ and $\simeq$ correspond to the antisymmetric and equivalence relations, respectively. For two random losses X, Y, the relation $X \succeq Y$ indicates that X is preferred over Y, or equivalently, that Y is considered more dangerous than X. A numerical representation of a preference $\succeq$ is a mapping $\rho : \mathcal{X} \to \mathbb{R}$, such that $X \succeq Y \iff \rho(X) \leqslant \rho(Y)$. Note that $\succeq$ can be represented by a mapping ρ if $\succeq$ is separable; see e.g., Drapeau and Kupper (2013).[6] We use the following axioms, where all random variables are tacitly assumed to be in L^p for some fixed $p \in [1, \infty]$.

A1 (Monotonicity). If $X_1 \leqslant X_2$, then $X_1 \succeq X_2$.

A2 (Translation-invariance). For any $c \in \mathbb{R}$, $X \succeq Y$ if and ony if $X + c \succeq Y + c$.

A3 (Weak positive homogeneity). If $\mathbb{E}[X] = \mathbb{E}[Y]$ and $X \succeq Y$, then $\lambda X \succeq \lambda Y$ for any $\lambda > 0$.

A4 (Risk aversion). If $X \leqslant_{\mathrm{cx}} Y$, then $X \succeq Y$. In addition, $\mathbb{E}[X] \succ X$ for any non-constant X.

A5 (Solvability). There exists $c \in \mathbb{R}$ such that $X \simeq c$.

A6 (Weak convexity). If $\mathbb{E}[X] = \mathbb{E}[Y]$ and $X \simeq Y$, then $\lambda X + (1 - \lambda)Y \succeq X$ for all $\lambda \in [0, 1]$.

A7 (Continuity). For every X, the sets $\{Y \in L^p : Y \succeq X\}$ and $\{Y \in L^p : X \succeq Y\}$ are L^p-closed.

These axioms are standard, and we refer to Yaari (1987), Drapeau and Kupper (2013), Föllmer and Schied (2016, Chapeter 2) and Grechuk et al. (2012) for interpretations and discussions of these axioms. The following result gives an axiomatization of MD_g^D in Definition 3.

Theorem 2. *Fix $p \in [1, \infty]$. A preference $\succeq$ on L^p satisfies Axioms A1–A7 if and only if $\succeq$ can be represented by $\mathrm{MD}_g^D = g \circ D + \mathbb{E}$ for some $D \in \overline{\mathcal{D}}^p$ and $g \in \mathcal{G}$ that is strictly increasing.*

Compared to the representation result of mean-deviation models in Grechuk et al. (2012), our stronger version of translation-invariance pins down the more explicit form of monotonic mean-deviation measures. We will also establish an explicit one-to-one correspondence between properties of the risk measure MD_g^D and properties of the risk-weighting function g in Section 4.2.

For the detailed differences between our axiomatization and that of Grechuk et al. (2012), see Appendix B. A subtle difference between Theorem 2 and Definition 3 is that g is strictly increasing in Theorem 2 but not necessarily so in Definition 3. An axiomatization of MD_g^D with g not necessarily strictly increasing is an open question, as we were unable to identify proper relaxations of the proposed axioms.

4.2 Characterizations of convex and coherent risk measures

We continue to study the properties of MD_g^D. Specifically, we characterize g such that MD_g^D belongs to the class of coherent risk measures or convex risk measures. Moreover, we consider *star-shaped risk measures*, which are monetary risk measures ρ further satisfying

[5] A preorder is a binary relation on $\mathcal{X}$, which is reflexive and transitive. A binary relation $\succeq$ is reflexive if $X \succeq X$ for all $X \in \mathcal{X}$, and transitive if $X \succeq Y$ and $Y \succeq Z$ imply $X \succeq Z$. A total preorder is a preorder which in addition is complete, that is, $X \succeq Y$ or $Y \succeq X$ for all $X, Y \in \mathcal{X}$.

[6] A total preorder $\succeq$ is separable if there exists a countable set $\mathcal{X} \subseteq L^p$ for $p \in [1, \infty]$ such that for any $x, y \in \mathcal{X}$ with $x \succ y$ there is $z \in \mathcal{X}$ for which $x \succeq z \succeq y$.

[SS] Star-shapedness: $\rho(0) = 0$ and $\rho(\lambda X) \leqslant \lambda\rho(X)$ for all $X \in \mathcal{X}$ and $\lambda \in [0, 1]$.

Similarly, a function $g : [0, \infty) \to \mathbb{R}$ is *star-shaped* if $g(0) = 0$ and $g(\lambda x) \leqslant \lambda g(x)$ for all $x \in [0, \infty)$ and $\lambda \in [0, 1]$. Star-shaped risk measures are characterized by Castagnoli et al. (2022) as the infimum of normalized (i.e., $\rho(0) = 0$) convex risk measures. Under normalization, star-shapedness is weaker than both convexity and positive homogeneity. We refer to Herdegen and Khan (2024), Laeven et al. (2024) and Nie et al. (2024) for more recent developments on star-shaped risk measures.

Theorem 3. *Suppose that $D \in \overline{\mathcal{D}}^p$ for $p \in [1, \infty]$ and $g \in \mathcal{G}$. The following statements hold.*

(i) MD_g^D *is a coherent risk measure if and only if g is linear.*

(ii) MD_g^D *is a convex risk measure if and only if g is convex.*

(iii) MD_g^D *is a star-shaped risk measure if and only if g is star-shaped.*

Proof. The sufficiency is straightforward. To show necessity, let X be such that $\mathbb{E}[X] = 0$ and $D(X) = 1$; such X exists due to Property (D3). The coherence of MD_g^D implies that for all $x > 0$,

$$g(x) = \mathrm{MD}_g^D(xX) = x\mathrm{MD}_g^D(X) = xg(1).$$

This implies that g is linear.

(ii) To determine sufficiency, if g is convex, then MD_g^D is a convex risk measure because the expectation is linear and D is convex. To show necessity, take $x, y \geqslant 0$ and $\lambda \in [0, 1]$. Let X be such that $\mathbb{E}[X] = 0$ and $D(X) = 1$. Since MD_g^D is convex and D satisfies (D3), we have

$$\begin{aligned} g(\lambda x + (1-\lambda)y) &= g \circ D((\lambda x + (1-\lambda)y)X) \\ &= \mathrm{MD}_g^D((\lambda x + (1-\lambda)y)X) \\ &\leqslant \lambda\mathrm{MD}_g^D(xX) + (1-\lambda)\mathrm{MD}_g^D(yX) = \lambda g(x) + (1-\lambda)g(y). \end{aligned}$$

Thus, g is convex.

(iii) To see sufficiency, if g is star-shaped, then MD_g^D is star-shaped because the expectation is linear and D satisfies (D3). Conversely, let X be such that $\mathbb{E}[X] = 0$ and $D(X) = 1$. For any $x \in [0, \infty)$ and $\lambda \in [0, 1]$, it follows from the star-shapedness of MD_g^D that $g(0) = \mathrm{MD}_g^D(0) = 0$ and

$$g(\lambda x) = \mathrm{MD}_g^D(\lambda xX) \leqslant \lambda\mathrm{MD}_g^D(xX) = \lambda g(x).$$

This implies that g is star-shaped. □

By Theorem 3 (i), MD_g^D is coherent if and only if

$$\mathrm{MD}_g^D(X) = \lambda D(X) + \mathbb{E}[X] = \lambda R(X) + (1-\lambda)\mathbb{E}[X], \quad X \in L^p$$

for some $\lambda \in [0, 1]$, where $R = D + \mathbb{E}$ is a coherent risk measure. In fact, positive homogeneity of MD_g^D is sufficient for g to be linear, as seen from the proof of (i). Therefore, positive homogeneity and coherence are equivalent for a monotonic mean-deviation measure. Moreover, following the same proof, the result in (ii) can be strengthened to a more general form without monotonicity: For any function $g : [0, \infty) \to \mathbb{R}$ and $D \in \mathcal{D}^p$ with $p \in [1, \infty]$, we have that MD_g^D is convex if and only if g is convex. As shown in Proposition 3 of Castagnoli et al. (2022), if MD_g^D is subadditive, then the three (coherent, convex, star-shaped) classes of risk measures in Theorem 3 (i)–(iii) coincide.

For the special choice of $D = \mathrm{ES}_\alpha - \mathbb{E}$ where $\alpha \in (0, 1)$, Han et al. (2023) obtained characterizations for MD_g^D to be coherent, convex, or consistent risk measures. Theorem 3 extends this result to deviation measures. In the following proposition, we obtain an alternative representation result for MD_g^D when g is convex.

Proposition 2. *Fix $p \in [1,\infty]$. For $g \in \mathcal{G}$ and $D \in \overline{\mathcal{D}}^p$, MD_g^D is a convex risk measure if and only if*

$$\mathrm{MD}_g^D(X) = \lambda \mathbb{E}[(D(X) - Y)_+] + \mathbb{E}[X]$$

for some non-negative random variable $Y \in L^1$ and some constant $\lambda \in [0,1]$. In particular, MD_g^D is a coherent risk measure if and only if $Y = 0$.

Proof. We need to show that g is an increasing, convex function which satisfies 1-Lipschitz continuity if and only if $g(x) = \lambda \mathbb{E}[(x - Y)_+]$ for some $Y \geqslant 0$ and $0 \leqslant \lambda \leqslant 1$. This is known in the literature; see Theorems 1 and 6 of Williamson (1956). □

By Theorem 1, we know that MD_g^D is a consistent risk measure, yet it fails to satisfy convexity when $g \in \mathcal{G}$ is not convex, as shown in Theorem 3. Thus, our results allow for explicit formulas for many consistent risk measures that are not convex, while existing examples of consistent but non-convex risk measures are often obtained by taking an infimum over convex risk measures. For instance, take $g(x) = \lambda \mathbb{E}[x \wedge Y]$ for some non-negative Y and $\lambda \in [0,1]$. It is obvious that g is concave and satisfies 1-Lipschitz continuity. In this case, $\mathrm{MD}_g^D(X)$ can be expressed as

$$\mathrm{MD}_g^D(X) = \lambda \mathbb{E}[D(X) \wedge Y] + \mathbb{E}[X],$$

which is a consistent but not convex risk measure. Furthermore, Theorem 3 implies that MD_g^D is a convex but not coherent risk measure if $g \in \mathcal{G}$ is convex yet non-linear. This insight opens up a new perspective for constructing risk measures within the class of monotone mean-deviation risk measures. Specifically, it guides us in developing risk measures that are consistent yet not convex, or alternatively, convex but not coherent, all while possessing an explicit formulation. By assuming that $g(x) = \mathbb{E}[(x - Y)_+]$ or $g(x) = \mathbb{E}[x \wedge Y]$ for some non-negative random variable Y, we can construct many convex or consistent risk measures with explicit forms which appear to be new in the literature.

Example 1. Let $g(x) = \mathbb{E}[(x - Y)_+]$ for some $Y \geqslant 0$ and $D \in \overline{\mathcal{D}}^p$ with some $p \in [1,\infty]$.

(i) If Y follows an exponential distribution with parameter $\beta > 0$, that is, $\mathbb{P}(Y > y) = e^{-\beta y}$, then $g(x) = x + \left(e^{-\beta x} - 1\right)/\beta$. According to Proposition 2, we have

$$\rho(X) = \mathbb{E}[X] + D(X) + \frac{1}{\beta}\left(e^{-\beta D(X)} - 1\right),$$

which is a convex risk measure.

(ii) If Y follows a Pareto distribution with tail parameter $\theta > 0$, that is, $\mathbb{P}(Y > y) = (1 + y)^{-\theta}$ for $y \geqslant 0$, then

$$g(x) = \begin{cases} x + \left((1+x)^{1-\theta} - 1\right)/(\theta - 1), & \theta \neq 1, \\ x - \log(1+x), & \theta = 1. \end{cases}$$

This yields

$$\rho(X) = \begin{cases} \mathbb{E}[X] + D(X) + \left((1+D(X))^{1-\theta} - 1\right)/(\theta - 1), & \theta \neq 1, \\ \mathbb{E}[X] + D(X) - \log(1 + D(X)), & \theta = 1, \end{cases}$$

and ρ is a convex risk measure.

Example 2. Let $g(x) = \mathbb{E}[x \wedge Y]$ for some $Y \geqslant 0$ and $D \in \overline{\mathcal{D}}^p$ for some $p \in [1, \infty]$.

(i) If Y follows an exponential distribution with parameter $\beta > 0$, then $g(x) = \left(1 - e^{-\beta x}\right)/\beta$. It follows that

$$\rho(X) = \mathbb{E}[X] + \frac{1}{\beta}\left(1 - e^{-\beta D(X)}\right),$$

which is a consistent risk measure but not a convex risk measure.

(ii) If Y follows a Pareto distribution with a tail parameter $\theta > 0$, then

$$g(x) = \begin{cases} \left(1 - (1+x)^{1-\theta}\right)/(\theta - 1), & \theta \neq 1, \\ \log(1+x), & \theta = 1. \end{cases}$$

This yields

$$\rho(X) = \begin{cases} \mathbb{E}[X] + \dfrac{1 - (1 + D(X))^{1-\theta}}{\theta - 1}, & \theta \neq 1, \\ \mathbb{E}[X] + \log(1 + D(X)), & \theta = 1, \end{cases}$$

which is a consistent risk measure but not a convex risk measure.

5 Dual representation

In this section, we investigate the dual representation of monotonic mean-deviation measures that are convex. Before presenting the main result, we introduce some notation. For $p \in [1, \infty)$, denote by q the conjugate dual of p, i.e., $q = (1 - 1/p)^{-1}$. Let $\mathcal{A}_p = \{Z \in L^q : Z \geqslant 0,\ \mathbb{E}[Z] = 1\}$. For a convex $g \in \mathcal{G}$, we use g^* to represent its conjugate function, i.e., $g^*(y) = \sup_{x \geqslant 0}\{xy - g(x)\}$. One can easily check that g^* is increasing, convex and lower semicontinuous. Note that $g : [0, \infty) \to \mathbb{R}$ is increasing and 1-Lipschitz continuous with $g(0) = 0$. Denote by $a = \lim_{x\to\infty} g'(x) \in [0, 1]$, and we have $g^*(y) = 0$ for $y \leqslant 0$ and $g^*(y) = \infty$ for $y > a$. Hence, $g(x) = g^{**}(x) = \sup_{y \in [0,a]}\{xy - g^*(y)\}$ holds for $x \geqslant 0$ (see e.g., Proposition A.6 of Föllmer and Schied (2016)).

From Definition 3 and Theorem 3, convex monotonic mean-deviation measure MD_g^D on L^p with $p \in [1, \infty)$ can be defined by the form MD_g^D with $D \in \overline{\mathcal{D}}^p$ and convex $g \in \mathcal{G}$. For $D \in \overline{\mathcal{D}}^p$, denote by $R = D + \mathbb{E}$, which is a finite coherent risk measure on L^p. Moreover, the following dual representation holds:

$$R(X) = D(X) + \mathbb{E}[X] = \max_{Z \in \mathcal{A}} \mathbb{E}[XZ], \quad X \in L^p \tag{11}$$

for some convex and weakly compact set $\mathcal{A} \subseteq \mathcal{A}_p$.

Theorem 4. *Fix $p \in [1, \infty)$. Suppose that $g \in \mathcal{G}$ is convex and $D \in \overline{\mathcal{D}}^p$ with the relation in (11). We have*

$$\mathrm{MD}_g^D(X) = \max_{Z \in \mathcal{A}} \left\{ a\mathbb{E}[XZ] - g^*\left(\frac{a}{\sup\{\lambda \in [1, \infty) : \lambda(Z - 1) + 1 \in \mathcal{A}\}}\right)\right\} + (1 - a)\mathbb{E}[X], \quad X \in L^p,$$

where $a = \lim_{x\to\infty} g'(x)$ is in $(0, 1]$.

Proof. Under the given conditions, MD_g^D is a convex risk measure. It follows from Theorem 2.11 of Kaina and Rüschendorf (2009) that it admits a dual representation:

$$\mathrm{MD}_g^D(X) = \max_{Z \in \mathcal{A}_p}\{\mathbb{E}[XZ] - \beta(Z)\}, \quad X \in L^p,$$

for some $\beta : L^q \to (-\infty, \infty]$ that is convex and lower semicontinuous, given by

$$\beta(Z) = \sup_{X \in L^p} \{\mathbb{E}[XZ] - \mathrm{MD}_g^D(X)\}, \quad Z \in L^q.$$

Note that $g \in \mathcal{G}$ is 1-Lipshitz continuous and non-constant. Combining with its convexity yields $a \in (0, 1]$.

Next, we aim to prove that β has an explicit representation:

$$\beta(Z) = \begin{cases} g^*\left(\frac{1}{\sup\{\lambda \in [1/a,\infty) : \lambda(Z-1)+1 \in \mathcal{A}\}}\right), & Z \in \mathcal{Z}, \\ \infty, & \text{otherwise}, \end{cases} \tag{12}$$

where $\mathcal{Z} = \{aY + 1 - a : Y \in \mathcal{A}\}$. For $Z \in L^q$, we have

$$\begin{aligned} \beta(Z) &= \sup_{X \in L^p} \{\mathbb{E}[XZ] - \mathrm{MD}_g^D(X)\} \\ &= \sup_{X \in L^p} \{\mathbb{E}[XZ] - \mathbb{E}[X] - g(D(X))\} \\ &= \sup_{X \in L^p} \inf_{y \in [0,a]} \{\mathbb{E}[XZ] - \mathbb{E}[X] - D(X)y + g^*(y)\}, \end{aligned} \tag{13}$$

where we have used $g(x) = \sup_{y \in [0,a]}\{xy - g^*(y)\}$ in the last step. The objective function of (13) is convex and lower semicontinuous in y for any fixed X since g^* is convex and lower semicontinuous, and it is concave in X for any fixed y. By a minimax theorem (see e.g., Theorem 2 of Fan (1953)), we have

$$\begin{aligned} \beta(Z) &= \inf_{y \in [0,a]} \sup_{X \in L^p} \{\mathbb{E}[XZ] - \mathbb{E}[X] - D(X)y + g^*(y)\} \\ &= \inf_{y \in [0,a]} \sup_{X \in L^p} \{\mathbb{E}[XZ] - \mathbb{E}[X] - (R(X) - \mathbb{E}[X])y + g^*(y)\} \\ &= \inf_{y \in [0,a]} \sup_{X \in L^p} \inf_{Y \in \mathcal{A}} \{\mathbb{E}[(Z - 1 + y - yY)X] + g^*(y)\}, \end{aligned} \tag{14}$$

where we have used (11) in the second and third steps. Obviously, the objective function of (14) is convex and continuous with respect to the weak topology in Y for any fixed X and concave in X. Additionally, $\mathcal{A}$ is convex and weakly compact. By the minimax theorem, we have

$$\beta(Z) = \inf_{y \in [0,a], Y \in \mathcal{A}} \sup_{X \in L^p} \{\mathbb{E}[(Z - 1 + y - yY)X] + g^*(y)\}. \tag{15}$$

Denote by $\widetilde{\mathcal{Z}} = \{yY + 1 - y : y \in [0, a],\ Y \in \mathcal{A}\}$. Note that the inner supremum problem above is infinite if $\mathbb{P}(Z - 1 + y - yY \neq 0) > 0$ and is equal to $g^*(y)$ if $Z - 1 + y - yY = 0$. We have that $\beta(Z) = \infty$ if $Z \in L^q \setminus \widetilde{\mathcal{Z}}$, and for $Z \in \widetilde{\mathcal{Z}}$, (15) reduces to

$$\begin{aligned} \beta(Z) &= \inf\{g^*(y) : y \in [0, a],\ Y \in \mathcal{A},\ y(Y - 1) + 1 = Z\} \\ &= \inf\left\{g^*(y) : y \in [0, a],\ \frac{Z - 1}{y} + 1 \in \mathcal{A}\right\} \\ &= \inf\left\{g^*\left(\frac{1}{\lambda}\right) : \lambda \in \left[\frac{1}{a}, \infty\right),\ \lambda(Z - 1) + 1 \in \mathcal{A}\right\} \\ &= g^*\left(\frac{1}{\sup\{\lambda \in [1/a, \infty) : \lambda(Z - 1) + 1 \in \mathcal{A}\}}\right), \end{aligned}$$

where the last equality holds because g^* is increasing. To verify (12), it remains to show $\mathcal{Z} = \widetilde{\mathcal{Z}}$, that is, $\{aY + 1 - a : Y \in \mathcal{A}\} = \{yY + 1 - y : y \in [0, a],\ Y \in \mathcal{A}\}$. It is clear that $\mathcal{Z} \subseteq \widetilde{\mathcal{Z}}$. Conversely, for any $Z \in \widetilde{\mathcal{Z}}$ with the representation $Z = yY + 1 - y$ for some $y \in [0, a]$ and $Y \in \mathcal{A}$, since $\mathcal{A}$ is convex and $1 \in \mathcal{A}$, we have that $\mathcal{Z}$ is convex and $1 \in \mathcal{Z}$. Note that $Z = (y/a)(aY + 1 - a) + (1 - y/a) \cdot 1$, where $y/a \in [0, 1]$ and $aY + 1 - a \in \mathcal{Z}$. It holds that $Z \in \mathcal{Z}$. This yields the converse direction. Hence, we have verified (12). Therefore, we have

$$\mathrm{MD}_g^D(X) = \max_{Z\in\mathcal{Z}} \left\{ \mathbb{E}[XZ] - g^*\left(\frac{1}{\sup\{\lambda \in [1/a, \infty) : \lambda(Z-1) + 1 \in \mathcal{A}\}} \right) \right\}, \quad X \in L^p,$$

where $\mathcal{Z} = \{aY + 1 - a : Y \in \mathcal{A}\}$. Moreover, for $Z \in \mathcal{Z}$ with the form $Z = aY + 1 - a$, where $Y \in \mathcal{A}$, it holds that

$$\begin{aligned}
&\mathbb{E}[XZ] - g^*\left(\frac{1}{\sup\{\lambda \in [1/a, \infty) : \lambda(Z-1) + 1 \in \mathcal{A}\}} \right) \\
&= \mathbb{E}[X(aY + 1 - a)] - g^*\left(\frac{1}{\sup\{\lambda \in [1/a, \infty) : \lambda a(Y-1) + 1 \in \mathcal{A}\}} \right) \\
&= a\mathbb{E}[XY] - g^*\left(\frac{a}{\sup\{\lambda \in [1, \infty) : \lambda(Y-1) + 1 \in \mathcal{A}\}} \right) + (1-a)\mathbb{E}[X].
\end{aligned}$$

This completes the proof. □

Below we give two specific examples of Theorem 4 by choosing the coherent risk measure R as ES or expectile (see e.g., Newey and Powell (1987) and Bellini et al. (2014)), which are popular in practice. This choice results in two classes of MD_g^D.

Example 3. Let $R = \mathrm{ES}_\alpha$ with $\alpha \in (0, 1)$, $D = R - \mathbb{E}$, $g \in \mathcal{G}$ be convex with $\lim_{x\to\infty} g'(x) = a$, and $\mathrm{MD}_g^D = g \circ D + \mathbb{E}$. The well-known dual representation of ES in Föllmer and Schied (2016, Example 4.40) gives $R(X) = \max_{Z\in\mathcal{A}} \mathbb{E}[XZ]$ for $X \in L^1$ where $\mathcal{A} = \{Z \in \mathcal{A}_\infty : Z \leqslant 1/(1-\alpha)\}$. Then

$$\begin{aligned}
\sup\{\lambda \in [1, \infty) : \lambda(Z-1) + 1 \in \mathcal{A}\} &= \sup\left\{ \lambda \in [1, \infty) : \lambda(\text{ess-sup}Z - 1) + 1 \leqslant \frac{1}{1-\alpha} \right\} \\
&= \frac{\alpha}{1-\alpha}(\text{ess-sup}Z - 1)^{-1}.
\end{aligned}$$

By Theorem 4, we obtain

$$\begin{aligned}
\mathrm{MD}_g^D(X) &= \max_{Z\in\mathcal{A}} \left\{ a\mathbb{E}[XZ] - g^*\left(\frac{(1-\alpha)a}{\alpha}(\text{ess-sup}Z - 1) \right) \right\} + (1-a)\mathbb{E}[X] \\
&= \sup_{\gamma\in[1,\frac{1}{1-\alpha}]} \sup\left\{ a\mathbb{E}[XZ] - g^*\left(\frac{(1-\alpha)(\gamma-1)a}{\alpha} \right) : Z \in \mathcal{A}_\infty,\ \text{ess-sup}Z = \gamma \right\} + (1-a)\mathbb{E}[X] \\
&= \sup_{\gamma\in[1,\frac{1}{1-\alpha}]} \left\{ a\mathrm{ES}_{1-\frac{1}{\gamma}}(X) - g^*\left(\frac{(1-\alpha)(\gamma-1)a}{\alpha} \right) \right\} + (1-a)\mathbb{E}[X] \\
&= \sup_{\gamma\in[0,\alpha]} \left\{ a\mathrm{ES}_\gamma(X) - g^*\left(\frac{1-\alpha}{\alpha}\frac{\gamma a}{1-\gamma} \right) \right\} + (1-a)\mathbb{E}[X].
\end{aligned}$$

Suppose now $a = 1$, and we define $f : [0, 1] \to (-\infty, \infty]$ as

$$f(\gamma) = \begin{cases} g^*\left(\frac{1-\alpha}{\alpha}\frac{\gamma}{1-\gamma}\right), & \gamma \in [0, \alpha], \\ \infty, & \gamma \in (\alpha, 1]. \end{cases}$$

Obviously, f is an increaing and convex function on $[0, 1]$ as g^* and $\gamma \mapsto \gamma/(1-\gamma)$ are both increasing and convex. It holds that

$$\mathrm{MD}_g^D(X) = \sup_{\gamma\in[0,1]} \{\mathrm{ES}_\gamma(X) - f(\gamma)\}.$$

A functional of the form $\sup_{\gamma\in[0,1]}\{\mathrm{ES}_\gamma(X) - h(\gamma)\}$ for a general function h is called an *adjusted Expected Shortfall* (AES) by Burzoni et al. (2022). Different from the general class of AES considered by Burzoni et al. (2022), the subclass MD_g^D has an explicit formula, i.e., $\mathrm{MD}_g^D(X) = g(\mathrm{ES}_\alpha(X)) + \mathbb{E}[X]$.

Example 4. An expectile at level $\alpha \in (0, 1)$, denoted by ex_α, is defined as the solution of the following equation:

$$\alpha\mathbb{E}[(X - x)_+] = (1 - \alpha)\mathbb{E}[(X - x)_-], \quad X \in L^1.$$

When $\alpha \geqslant 1/2$, ex_α is a convex risk measure admitted a dual representation (see e.g., Proposition 8 of Bellini et al. (2014)):

$$\mathrm{ex}_\alpha(X) = \max_{Z\in\mathcal{A}} \mathbb{E}[XZ] \quad \text{with } \mathcal{A} = \left\{Z \in \mathcal{A}_\infty : \frac{\text{ess-sup}Z}{\text{ess-inf}Z} \leqslant \frac{\alpha}{1-\alpha}\right\}.$$

Let $R = \mathrm{ex}_\alpha$ with $\alpha \in [1/2, 1)$, $D = R - \mathbb{E}$ and $g \in \mathcal{G}$ be convex with $\lim_{x\to\infty} g'(x) = a$, and let $\mathrm{MD}_g^D = g \circ D + \mathbb{E}$. It holds that

$$\begin{aligned} \sup\{\lambda \in [1, \infty) : \lambda(Z - 1) + 1 \in \mathcal{A}\} &= \sup\left\{\lambda \in [1, \infty) : \frac{\lambda(\text{ess-sup}Z - 1) + 1}{\lambda(\text{ess-inf}Z - 1) + 1} \leqslant \frac{\alpha}{1-\alpha}\right\} \\ &= \frac{2\alpha - 1}{2\alpha - 1 + (1 - \alpha)\text{ess-sup}Z - \alpha\text{ess-inf}Z}. \end{aligned}$$

By Theorem 4, we obtain

$$\mathrm{MD}_g^D(X) = \sup_{Z\in\mathcal{A}} \left\{a\mathbb{E}[XZ] - g^*\left(\frac{a((1-\alpha)\text{ess-sup}Z - \alpha\text{ess-inf}Z)}{2\alpha - 1} + a\right)\right\} + (1 - a)\mathbb{E}[X].$$

Recalling the representation of a convex monotonic mean-deviation measure MD_g^D in Theorem 4, the smallest coherent risk measure (denoted by ρ_S) that dominates MD_g^D can be directly derived:

$$\rho_S(X) = \max_{Z\in\mathcal{A}} a\mathbb{E}[XZ] + (1 - a)\mathbb{E}[X] = aR(X) + (1 - a)\mathbb{E}[X] = aD(X) + \mathbb{E}[X], \quad X \in L^p.$$

Below we provide an analogous result for a broader class of functionals with the form of MD_g^D where $g \in \mathcal{G}$ is not necessarily convex, implying that MD_g^D may not be a convex risk measure.

Proposition 3. *Let $g \in \mathcal{G}$ and $D \in \overline{\mathcal{D}}^p$. The smallest coherent risk measure that dominates MD_g^D is $(\sup_{x>0} g(x)/x)D(X) + \mathbb{E}[X]$.*

Proof. The smallest positive homogeneous functional that dominates MD_g^D is given by

$$\rho(X) = \sup_{\lambda>0} \frac{\mathrm{MD}_g^D(\lambda X)}{\lambda} = \sup_{\lambda>0} \frac{g(\lambda D(X))}{\lambda} + \mathbb{E}[X]$$
$$= D(X) \sup_{\lambda>0} \frac{g(\lambda)}{\lambda} + \mathbb{E}[X].$$

Hence, we obtain the desired result. □

Proposition 3 addresses the case where g is convex, which aligns with the observation in Theorem 4. This is because $\sup_{x>0} g(x)/x = \lim_{x\to\infty} g'(x)$ for convex $g \in \mathcal{G}$. Despite the simplicity of the proof of above proposition, the smallest dominating coherent risk measure of a given risk measure has several interesting applications; see Wang et al. (2015) in the context of subadditivity, and Herdegen and Khan (2024) in the context of arbitrage induced by risk measure.

6 Non-parametric estimation

In this section, we first introduce the definition of signed Choquet integrals and then examine the properties of non-parametric estimators for MD_g^D, where D is chosen as a signed Choquet integral.

Define

$$\mathcal{H} = \{h : \ h \text{ maps from } [0,1] \text{ to } \mathbb{R} \text{ is of bounded variation with } h(0) = h(1) = 0\}.$$

The elements in $\mathcal{H}$ are called distortion functions. A *signed Choquet integral* with the distortion function $h \in \mathcal{H}$ is a functional, denoted by D_h, that has the representation:

$$D_h(X) = \int_{\mathbb{R}} h\left(\mathbb{P}(X > x)\right) \mathrm{d}x. \tag{16}$$

The class of signed Choquet integrals has been characterized by Wang et al. (2020a,b) via comonotonic additivity.[7] A signed Choquet integral with the form (16) always satisfies (D3) in Definition 1. It also follows that D_h with $h \in \mathcal{H}$ satisfies (D2) since $\int_{\mathbb{R}} h(\mathbb{P}(X + c > x))\mathrm{d}x = \int_{\mathbb{R}} h(\mathbb{P}(X > x))\mathrm{d}x$ for all $c \in \mathbb{R}$. To derive a class of deviation measures on L^p from all signed Choquet integrals, we need to shrink the set $\mathcal{H}$ so that D_h additionally satisfies (D1) and (D4), as well as ensures finiteness on L^p. This leads to consider the following subset of $\mathcal{H}$:

$$\Phi^p = \{h \in \mathcal{H} : \ h \text{ is concave and } \|h'\|_q < \infty\}, \quad p \in [1,\infty),$$

where h' is the left derivative of h, $q = (1 - 1/p)^{-1}$, and $\|h'\|_q = (\int_0^1 |h'(t)|^q \,\mathrm{d}t)^{1/q}$ for $q \in [1,\infty)$ and $\|h'\|_\infty = \sup_{t\in(0,1)} |h'(t)|$. For $h \in \Phi^p$, D_h can be reformulated as

$$D_h(X) = \int_{\mathbb{R}} h\left(\mathbb{P}(X > x)\right) \mathrm{d}x = \int_0^1 \mathrm{VaR}_\alpha(X) h'(1-\alpha)\mathrm{d}\alpha. \tag{17}$$

Indeed, it follows from Theorem 3 of Wang et al. (2020b) that D_h satisfies (D4) if and only if $h \in \mathcal{H}$ is concave, thereby implying the property (D4) for D_h when $h \in \Phi^p$. The property (D1) of

[7]Random variables X and Y are said to be comonotonic if there exists $\Omega_0 \in \mathcal{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that $\omega, \omega' \in \Omega_0$ $(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0$. For a functional $\rho : \mathcal{X} \to \mathbb{R}$, we say that ρ is comonotonic additive, if for any comonotonic random variables $X, Y \in \mathcal{X}$, $\rho(X + Y) = \rho(X) + \rho(Y)$.

D_h with $h \in \Phi^p$ is obtained from the non-negativity of h. Moreover, Proposition 1 of Wang et al. (2020a) shows that $h \in \Phi^p$ is a sufficient condition for the finiteness of D_h on L^p. Therefore, we have concluded that D_h is a deviation measure in the sense of Definition 1 whenever $h \in \Phi^p$.

For $p \geqslant 1$, $h \in \Phi^p$ and $D_h : L^p \to \mathbb{R}$ defined by (17), one can observe that $\lambda D_h + \mathbb{E}$ satisfies [CA], [PH] and [SA] for any $\lambda \geqslant 0$. Combining Proposition 2 (ii) of Wang et al. (2020a) and Proposition 1, it is established that D_h is range-normalized if and only if $t \mapsto h(t) + t$ is increasing on $[0, 1]$, and $t \mapsto \lambda h(t) + t$ is not an increasing function on $[0, 1]$ for any $\lambda > 1$, which is equivalent to $h'(1) = -1$; we do not assume this condition in this section.

We now examine the properties of non-parametric estimators for MD_g^D, where $g \in \mathcal{G}$ and $D = D_h$ with $h \in \Phi^p$. These estimators can be derived from those of D, VaR, and the expectation, as detailed in this section. Suppose that $X_1, \dots, X_n$ are an iid sample from (the distribution of) a random variable $X \in L^p$. Recall that the empirical distribution $\widehat{F}_n$ of $X_1, \dots, X_n$ is given by

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{\{X_j \leqslant x\}}, \quad x \in \mathbb{R}.$$

Let $\widehat{\mathrm{MD}}_g^D(n)$ be the empirical estimator of $\mathrm{MD}_g^D(X)$, obtained by applying MD_g^D to the empirical distribution of $X_1, \dots, X_n$. We will establish the consistency and asymptotic normality of the empirical estimators, based on corresponding results on empirical estimators of $\mathbb{E}[X]$ and $D(X)$. Let $\widehat{x}_n$ and $\widehat{D}(n)$ be the empirical estimators of $\mathbb{E}[X]$ and $D(X)$ based on the first n sample data points. We make following standard regularity assumption on the distribution of the random variable X.

Assumption 1. The distribution F of X is supported on a convex set and has a positive density function f on the support. Denote by $\tilde{f} = f \circ F^{-1}$.

The proof of Theorem 5 below relies on standard techniques in empirical quantile processes, and it is given in Appendix C. In what follows, g' is the left derivative of g.

Theorem 5. *Fix $p \in [1, \infty)$. Let $g \in \mathcal{G}$ and $D = D_h$ for some $h \in \Phi^p$. Suppose that $X_1, \dots, X_n \in L^p$ are an iid sample from $X \in L^p$ and Assumption 1 holds. Then, $g(\widehat{D}(n)) + \widehat{x}_n \xrightarrow{\mathbb{P}} g(D(X)) + \mathbb{E}[X]$ as $n \to \infty$. Moreover, if $p < 2$ and $X \in L^\gamma$ for some $\gamma > 2p/(2-p)$, then we have*

$$\sqrt{n} \left(\widehat{\mathrm{MD}}_g^D(n) - \mathrm{MD}_g^D(X) \right) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \sigma_g^2\right),$$

in which

$$\sigma_g^2 = \int_0^1 \int_0^1 \frac{(h'(1-s)g'(D(X)) + 1)(h'(1-t)g'(D(X)) + 1)(s \wedge t - st)}{\tilde{f}(s)\tilde{f}(t)} \mathrm{d}t\mathrm{d}s. \tag{18}$$

The integrability condition $X \in L^\gamma$ with $\gamma > 2p/(2-p)$, required for asymptotic normality in Theorem 5, coincides with that in Jones and Zitikis (2003), where the authors established the asymptotic normality of empirical estimators for distortion risk measures. In particular, in case $p = 1$, we require $X \in L^\gamma$ with $\gamma > 2$, which is a common assumption in weighted empirical quantile processes without distortion; see e.g., Shao and Yu (1996). The condition $p < 2$ is also important. If $D_h \notin \Phi^2$, then D_h is not even finite on L^2, and we do not expect asymptotic normality in this case.

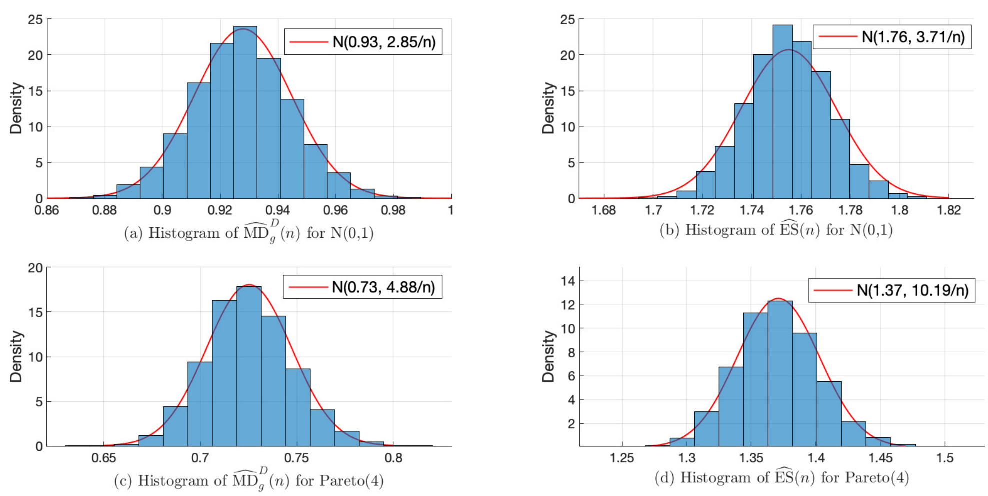
Note that the asymptotic variance σ_g in (18) is decreasing in the left derivative of g. Therefore, if we replace g by $\tilde{g} \in \mathcal{G}$ satisfying $\tilde{g}' \leqslant g'$, then the asymptotic variance, and thus the estimation error, will decrease. Note that a larger g' corresponds to a larger sensitivity to risk, as it measures how MD_g^D changes when D increases. Therefore, Theorem 5 gives a trade-off between risk sensitivity and statistical efficiency.

In what follows, we present some simulation results based on Theorem 5. We assume that $g(x) = x + e^{-x} - 1$ and $g(x) = 1 - e^{-x}$, respectively. The simulation results are presented in the case of standard normal and Pareto risks with tail index 4. Let the sample size be $n = 10000$, and we repeat the procedure 5000 times.

First, let $h(t) = 1-t-(1-\alpha-t)_+/(1-\alpha)$ for $t \in [0,1]$ with $\alpha = 0.9$. It is straightforward to see $h \in \Phi^1$. Further, we set $D = D_h = \mathrm{ES}_\alpha - \mathbb{E}$, and $\mathrm{MD}_g^D(X)$ is given as $\mathrm{MD}_g^D(X) = g(\mathrm{ES}_\alpha(X) - \mathbb{E}[X]) + \mathbb{E}[X]$. In this case, we have $h'(1-t) = \mathbb{1}_{\{t \geqslant \alpha\}}/(1-\alpha) - 1$ and σ_g^2 in (18) can be computed explicitly. We compare the asymptotic variance of MD_g^D with that of ES_α, given by, via (18),

$$\sigma_{\mathrm{ES}}^2 = \frac{1}{(1-\alpha)^2} \int_\alpha^1 \int_\alpha^1 \frac{s \wedge t - st}{\tilde{f}(s)\tilde{f}(t)} \mathrm{d}t\mathrm{d}s.$$

In Figure 2 (a) and (b), the sample is simulated from the standard normal distribution. We can observe that, for $g(x) = x + e^{-x} - 1$ and $D = \mathrm{ES}_\alpha - \mathbb{E}$, empirical estimates of MD_g^D match quite well with the density function of $\mathrm{N}(0.93, 2.85/n)$. In contrast, the ES_α empirical estimates match with the density function of $\mathrm{N}(1.76, 3.71/n)$, whose asymptotic variance is larger than that of MD_g^D. In Figure 2 (c) and (d), the sample is simulated from the Pareto distribution with a tail index 4. We can observe that the $\mathrm{MD}_g^D(X)$ empirical estimates match quite well with the density function of $\mathrm{N}(0.73, 4.88/n)$ and the ES empirical estimates match with the density function of $\mathrm{N}(1.37, 10.19/n)$, whose asymptotic variance is also larger than the one of MD_g^D. Since g satisfies 1-Lipschitz continuity, the volatility of D is reduced via the distortion by g.



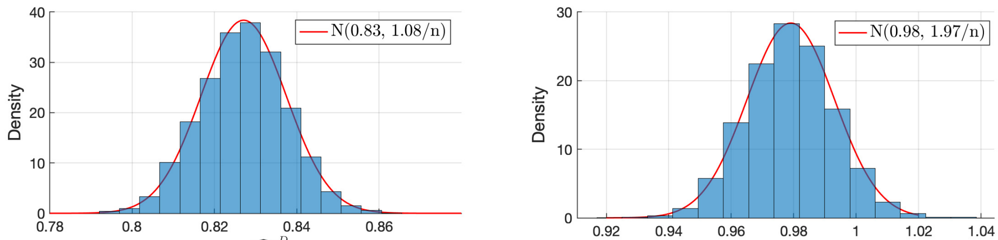
(a) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for N(0,1) (b) Histogram of $\widehat{\mathrm{ES}}(n)$ for N(0,1)

(c) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for Pareto(4) (d) Histogram of $\widehat{\mathrm{ES}}(n)$ for Pareto(4)

Figure 2: Left: $\widehat{\mathrm{MD}}_g^D(n)$ with $D = \mathrm{ES}_\alpha - \mathbb{E}$ and $g(x) = x + e^{-x} - 1$; Right: $\widehat{\mathrm{ES}}_\alpha(n)$

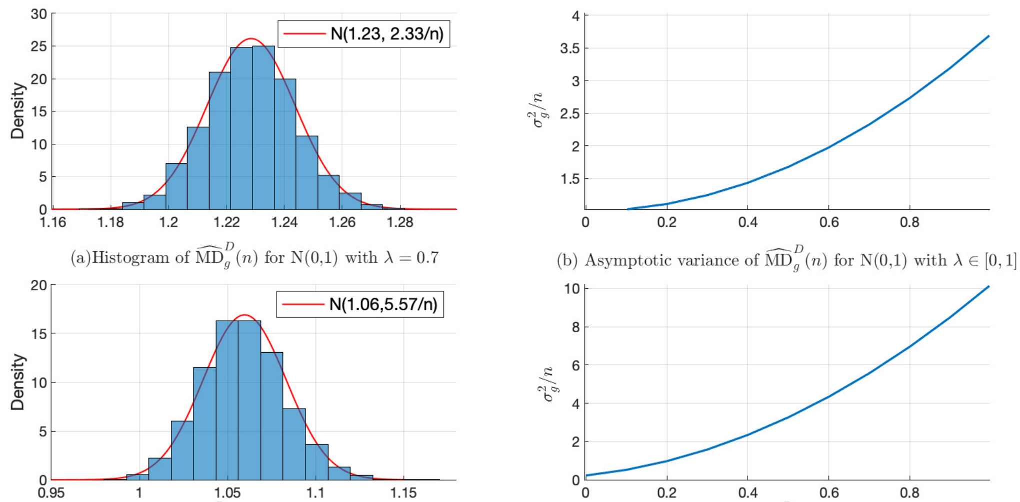
In Figure 3 (a) and (b), for $g(x) = 1 - e^{-x}$ and $D = \mathrm{ES}_\alpha - \mathbb{E}$, we can observe that the MD_g^D empirical estimates match quite well with the density functions of $\mathrm{N}(0.83, 1.08/n)$ and $\mathrm{N}(0.98, 1.97/n)$ when the samples are also simulated from the standard normal distribution or the Pareto distribution with a tail index 4. Moreover, the asymptotic variance in both cases is smaller than those of the ES empirical estimates.

If $g(x) = \lambda x$ with $\lambda \in (0,1)$, then $\mathrm{MD}_g^D(X) = \lambda \mathrm{ES}_\alpha(X) + (1-\lambda)\mathbb{E}[X]$. It is obvious that the asymptotic variance of $\mathbb{E}$/ES-mixture is increasing in λ and thus it is smaller than that of ES. Moreover, if $\lambda = 1$, then $\mathrm{MD}_g^D = \mathrm{ES}_\alpha$, and the values of σ_g^2/n in Figure 4 (b) and (d) equal to those in Figure 2 (b) and (d).



(a) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for N(0,1)   (b) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for Pareto(4)

Figure 3: $\widehat{\mathrm{MD}}_g^D(n)$ with $D = \mathrm{ES}_\alpha - \mathbb{E}$ and $g(x) = 1 - e^{-x}$



(a)Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for N(0,1) with $\lambda = 0.7$

(b) Asymptotic variance of $\widehat{\mathrm{MD}}_g^D(n)$ for N(0,1) with $\lambda \in [0,1]$

(c) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for Pareto(4) with $\lambda = 0.7$

(d) Asymptotic variance of $\widehat{\mathrm{MD}}_g^D(n)$ for Pareto(4) with $\lambda \in [0,1]$

Figure 4: $\widehat{\mathrm{MD}}_g^D(n)$ with $D = \mathrm{ES}_\alpha - \mathbb{E}$ and $g(x) = \lambda x$

Below we give another example of D. For $X \in L^1$, let X_1, X_2, X be iid, and

$$D(X) = \mathrm{Gini}(X) := \frac{1}{2}\mathbb{E}\left[|X_1 - X_2|\right]. \tag{19}$$

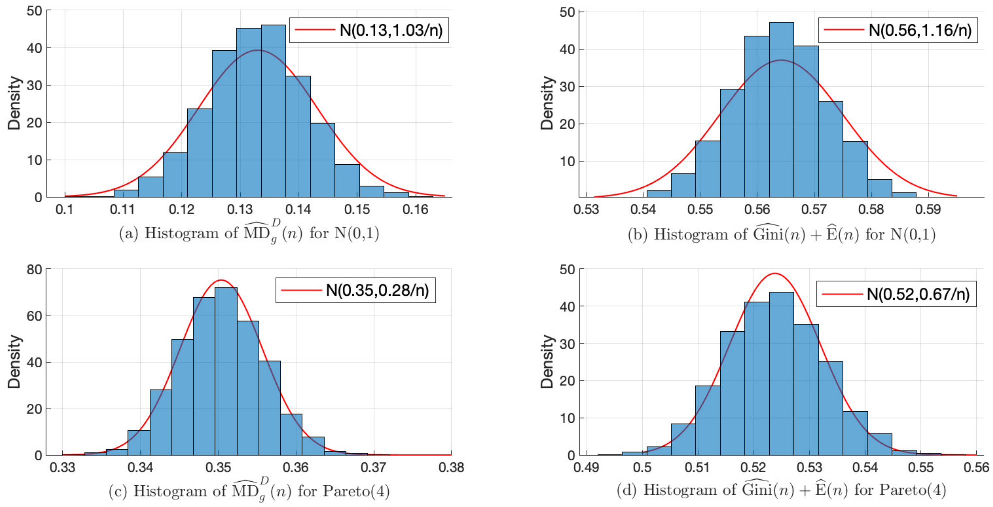
The Gini deviation is a signed Choquet integral with a distortion function $h \in \Phi^1$ given by $h(t) = t - t^2$ for $t \in [0,1]$ (see e.g., Denneberg (1990)), i.e., $D = \mathrm{Gini} = D_h$. Then we have

$$\sigma_g^2 = \int_0^1 \int_0^1 \frac{((2s-1)g'(\mathrm{Gini}(X)) + 1)((2t-1)g'(\mathrm{Gini}(X)) + 1)(s \wedge t - st)}{\tilde{f}(s)\tilde{f}(t)} \mathrm{d}t\mathrm{d}s.$$

Note that the asymptotic variance for estimating $\mathrm{Gini}(X) + \mathbb{E}[X]$, denoted by $\sigma^2_{\mathrm{Gini}+\mathbb{E}}$, equals

$$\sigma^2_{\mathrm{Gini}+\mathbb{E}} = \int_0^1 \int_0^1 \frac{4ts(s \wedge t - st)}{\tilde{f}(s)\tilde{f}(t)} \mathrm{d}t\mathrm{d}s.$$

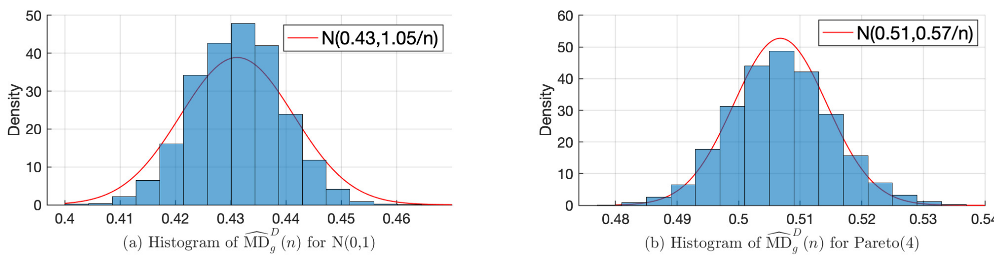
The simulation results are presented in Figures 5 and 6 for $D = \mathrm{Gini}$ in the case of the standard normal distribution and the Pareto distribution with tail index 4, which also confirm the asymptotic normality of the empirical estimators in Theorem 5. Similarly, the asymptotic variance of $\mathbb{E} + \mathrm{Gini}$ is larger than the one of MD_g^D based on $D = \mathrm{Gini}$.



(a) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for N(0,1) (b) Histogram of $\widehat{\mathrm{Gini}}(n) + \widehat{\mathbb{E}}(n)$ for N(0,1)

(c) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for Pareto(4) (d) Histogram of $\widehat{\mathrm{Gini}}(n) + \widehat{\mathbb{E}}(n)$ for Pareto(4)

Figure 5: Left: $\widehat{\mathrm{MD}}_g^D(n)$ with $g(x) = x + e^{-x} - 1$ and $D = \mathrm{Gini}$; Right: $\widehat{\mathrm{Gini}}(n) + \widehat{\mathbb{E}}(n)$



(a) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for N(0,1) (b) Histogram of $\widehat{\mathrm{MD}}_g^D(n)$ for Pareto(4)

Figure 6: $\widehat{\mathrm{MD}}_g^D(n)$ with $D = \mathrm{Gini}$ and $g(x) = 1 - e^{-x}$

7 Application to portfolio selection

In this section, we consider portfolio selection problems based on MD_g^D. Let a random vector $\mathbf{X} \in \mathcal{X}^n$ represent log-losses (i.e., the negation of log-return of the daily asset prices; see McNeil et al. (2015)) from n assets and a vector $\mathbf{w} = (w_1, \dots, w_n) \in \Delta_n$ of portfolio weights, where Δ_n is the standard n-simplex, given by

$$\Delta_n = \{(w_1, \dots, w_n) \in [0,1]^n : w_1 + \cdots + w_n = 1\}.$$

The total loss of the portfolio is $\mathbf{w}^\top \mathbf{X}$, and the optimization problem is formulated as

$$\min_{\mathbf{w} \in \Delta_n} \mathrm{MD}_g^D(\mathbf{w}^\top \mathbf{X}). \tag{20}$$

Note that the convexity of g implies that MD_g^D is a convex risk measure (see Theorem 3), and in this case, problem (20) is a convex optimization problem.

We select the 4 largest stocks from each of the 10 different sectors of S&P 500, ranked by market cap at the beginning of 2014, as the portfolio compositions (40 stocks in total). The historical asset prices are collected from Yahoo Finance, covering the period from January 2, 2014, to December 29, 2023, with a total of 2516 observations of daily losses. We use the first two years' data for training and start investment strategies at the beginning of 2016. The initial wealth is set to 1, and the risk-free rate is $r = 2.13\%$, which is the 10-year yield of the US treasury bill in Jan 2016. Note that the risk-free asset is not used to construct portfolios, but only used to calculate the Sharpe ratios.

Our portfolio strategies rebalance at the beginning of each month by solving (20) and we assume no transaction cost. For each period, we use an empirical distribution of the previous 500 log-loss data to estimate the risk measure, i.e., using an empirical estimator as described in Section 6. This is the simplest method of computing the risk measure, although the standard method in practice is to fit a time-series model. We choose this simple method for illustrative purposes.

We consider MD_g^D by choosing $D = \mathrm{ES}_\alpha - \mathbb{E}$ with $\alpha = 0.9$ and varying g, since our main novelty lies in the risk-weighting function g. In particular, we consider the following class of convex functions g_β indexed by a parameter $\beta > 0$ as in Example 1 (i), given by

$$g_\beta(x) = x + \frac{1}{\beta}\left(e^{-\beta x} - 1\right).$$

The parameter β has a natural interpretation of describing the convexity of g_β; that is, a smaller β means a more convex g_β. This is because g''_β/g'_β is decreasing in β (see Ross (1981) for comparing convexity of functions). Note that $g_\beta(x) \to x$ as $\beta \to \infty$, which represents a linear risk-weighting function.

At each period, the problem is to minimize MD_g^D over $\mathbf{w} \in \Delta_n$, that is,

$$\min_{\mathbf{w}\in\Delta_n} : \quad \mathbb{E}[\mathbf{w}^\top \mathbf{X}] + g_\beta(\mathrm{ES}_\alpha(\mathbf{w}^\top \mathbf{X}) - \mathbb{E}[\mathbf{w}^\top \mathbf{X}]),$$

where $\mathbf{X}$ follows the empirical distribution of the log-loss vector of the previous 500 trading days. By using the ES optimization formula of Rockafellar and Uryasev (2002), that is,

$$\mathrm{ES}_\alpha(X) = \min_{x\in\mathbb{R}} \left\{ x + \frac{1}{1-\alpha}\mathbb{E}[(X-x)_+] \right\}, \quad X \in L^1,$$

we can write the MD_g^D minimization problem as

$$\min_{\mathbf{w}\in\Delta_n,\ x\in\mathbb{R}} : \quad \mathbb{E}[\mathbf{w}^\top \mathbf{X}] + g_\beta\left(x + \frac{1}{1-\alpha}\mathbb{E}[(\mathbf{w}^\top \mathbf{X} - x)_+] - \mathbb{E}[\mathbf{w}^\top \mathbf{X}] \right). \tag{21}$$

The problem (21) is jointly convex in $\mathbf{w}$ and x and therefore can be easily solved numerically via modern computational programs such as MATLAB.

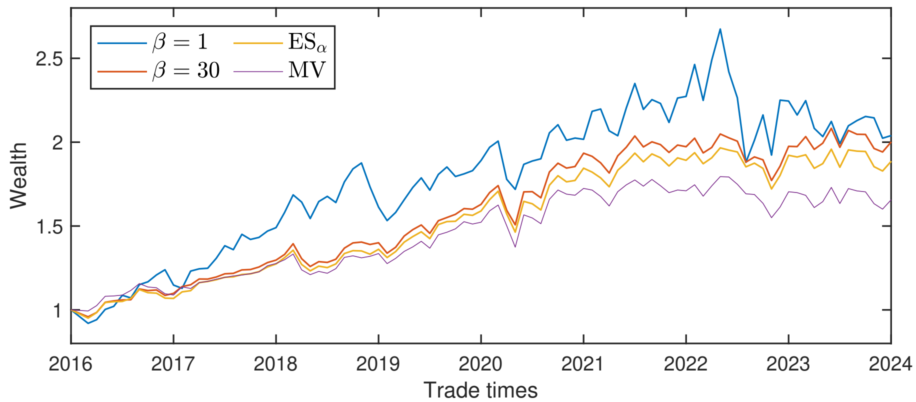
We choose $\beta = 1, 3, 10, 30, 100$ to study the effect of β. We compare them with a portfolio that simply minimizes ES_α (corresponding to $\beta = \infty$) and a Markowitz (1952) portfolio (by fixing an expected log-return at 10% and minimizing the variance at each period). The portfolio performance is reported in Figure 7. Summary statistics, including the annualized return (AR), the annualized volatility (AV), and the Sharpe ratio (SR) are reported in Table 1.

Table 1: Annualized return (AR), annualized volatility (AV), and Sharpe ratio (SR) for different portfolio strategies from Jan 2016 to Dec 2023

%	$\beta = 1$	$\beta = 3$	$\beta = 10$	$\beta = 30$	$\beta = 100$	ES_α	MV
AR	9.28	9.05	9.23	9.14	8.71	8.41	6.77
AV	19.92	15.21	12.78	12.26	12.05	12.10	11.47
SR	35.89	45.47	55.54	57.19	54.59	51.95	38.76

Our findings suggest that MD_g^D minimizing portfolios have some intuitive features. A smaller β,

Figure 7: Wealth processes for portfolios, 40 stocks, Jan 2016 - Dec 2023



meaning a more convex and smaller risk-weighting function g_β, roughly leads to a larger annualized return and a larger annualized volatility. This is intuitive because the weight on the deviation measure $\mathrm{ES}_\alpha - \mathbb{E}$ is smaller for smaller β, thus putting a higher value on the return. On the other hand, a large Sharpe ratio is attained around $\beta = 30$, suggesting that a suitable level of β can balance the return and the volatility quite well. The more convex g_β is, the more MD neglects small values of deviation. This may partially explain the observation in Figure 7, where the curve corresponding to $\beta = 1$ has many more fluctuations than the one corresponding to $\beta = 30$. We admit that this observation does not have a theoretical justification, and it is based only on one dataset, so we do not intend to generalize. To fully understand the effect of the convexity of g on portfolio selection, future studies are needed.

Of course, our objective is not to identify which strategy yields the highest return or Sharpe ratio in financial practice, a question that depends highly on the market situation and economic environment. Instead, the empirical results presented here mainly illustrate the interpretability of the strategy for portfolio selection based on MD_g^D. Moreover, the optimization and portfolio strategies are easy to implement.

8 Conclusion

Even though mean-deviation measures are widely considered in the literature and have a lot of attractive features, there are few systemic treatments in the literature. In this paper, we studied the class MD_g^D of mean-deviation measures whose form is a combination of the deviation-related functional and the expectation, which enriches the axiomatic theory of risk measures. In particular, the obtained class always belongs to the class of consistent risk measures. We showed that MD_g^D can be coherent, convex or star-shaped risk measures, identified with the corresponding properties of the risk-weighting function g. By looking at this new class, the gap between convex risk measures and consistent risk measures, arguably opaque in the literature due to lack of explicit examples, becomes transparent. The empirical estimators of MD_g^D can be formulated when D is chosen as a signed Choquet integral, and the asymptotic normality of the estimators is established. We find that the asymptotic variance of MD_g^D is smaller than the one of risk measures without distortion on deviation; a useful feature in statistical estimation. This intuitively illustrates a trade-off between statistical efficiency and sensitivity to risk.

We discuss several future directions for the research of MD_g^D. In fact, the form of MD_g^D (not necessarily monotonic) includes many commonly used reinsurance premium principles as special cases; see, e.g., the variance related principles (Furman and Landsman (2006) and Chi (2012)) and the Denneberg's absolute deviation principle (Tan et al. (2020)). Thus, it would be interesting to formulate the optimal reinsurance problem where the reinsurance principle is computed by MD_g^D. It is also meaningful to consider risk sharing problems and portfolio selection problems under the criterion of minimizing MD_g^D, following a similar framework to that of Grechuk et al. (2012, 2013) and Grechuk and Zabarankin (2012). Another direction of generalization is to relax cash-additivity we imposed throughout the paper to cash-subadditivity, as this allows for non-constant eligible assets when computing regulatory capital requirement; see El Karoui and Ravanelli (2009) and Farkas et al. (2014). Finally, we worked throughout with law-invariant mean-deviation measures with respect to a fixed probability measure. When the probability measure is uncertain, one needs to develop a framework of mean-deviation measures that can incorporate uncertainty and multiple scenarios in some forms (e.g., Cambou and Filipović (2017), Delage et al. (2019) and Fadina et al. (2023)).

Acknowledgements. We thank the Editor, an Associate Editor, two anonymous referees for helpful comments. Xia Han is supported by the National Natural Science Foundation of China (Grant Nos. 12301604, 12371471, 12471449). Ruodu Wang is supported by the Natural Sciences and Engineering Research Council of Canada (CRC-2022-00141, RGPIN-2024-03728) and the Sun Life Research Fellowship.

References

Alcantud, J. C. R. and Bosi, G. (2003). On the existence of certainty equivalents of various relevant types. *Journal of Applied Mathematics*, **2003**(9), 447–458.

Artzner, P., Delbaen, F., Eber, J.-M. and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(3), 203–228.

Bellini, F., Fadina, T., Wang, R. and Wei, Y. (2022). Parametric measures of variability induced by risk measures. *Insurance: Mathematics and Economics*, **106**, 270–284.

Bellini, F., Klar, B., Müeller, A. and Gianin, E. R. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **54**, 41–48.

Bernard, C., Pesenti, S. M. and Vanduffel, S. (2024). Robust distortion risk measures. *Mathematical Finance*, **34**(3), 774–818.

Burzoni, M., Munari, C. and Wang, R. (2022). Adjusted expected shortfall. *Journal of Banking and Finance*, **134**, 106297.

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

Chi, Y. (2012). Optimal reinsurance under variance related premium principles. *Insurance: Mathematics and Economics*, **51**(2), 310–321.

Cambou, M. and Filipović, D. (2017). Model uncertainty and scenario aggregation. *Mathematical Finance*, **27**(2), 534–567.

Dana R. A. (2005). A representation result for concave Schur-concave functions. *Mathematical Finance*, **15**(4), 613–634.

Del Barrio, E., Giné, E. and Utzet, F. (2005). Asymptotics for L_2 functionals of the empirical quantile process, with applications to tests of fit based on weighted Wasserstein distances. *Bernoulli*, **11**(1), 131–189.

Delage, E., Kuhn, D. and Wiesemann, W. (2019). "Dice"-sion-making under uncertainty: When can a random decision reduce risk? *Management Science*, **65**(7), 3282–3301.

Denneberg, D. (1990). Premium calculation: Why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin*, **20**, 181–190.

Dhaene, J., Vanduffel, S., Goovaerts, M. J., Kaas, R., Tang, Q. and Vyncke, D. (2006). Risk measures and comonotonicity: A review. *Stochastic Models*, **22**(4), 573–606.

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

El Karoui, N. and Ravanelli, C. (2009). Cash subadditive risk measures and interest rate ambiguity. *Mathematical Finance*, **19**(4), 562–590.

Esfahani, P. M. and Kuhn, D. (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, **171**(1-2), 115–166.

Fadina, T., Liu, Y. and Wang, R. (2023). A framework for measures of risk under uncertainty. *Finance and Stochastics*, forthcoming.

Fan, K. (1953). Minimax theorems. *Proceedings of the National Academy of Sciences*, **39**(1), 42–47.

Farkas, W., Koch-Medina, P. and Munari, C. (2014). Beyond cash-additive risk measures: When changing the numeraire fails. *Finance and Stochastics*, **18**, 145–173.

Filipović, D. and Svindland, G. (2008). Optimal capital and risk allocations for law- and cash-invariant convex functions. *Finance and Stochastics*, **12**, 423–439.

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. An Introduction in Discrete Time.* Fourth Edition. Walter de Gruyter, Berlin.

Frittelli, M. and Rosazza Gianin, E. (2002). Putting order in risk measures. *Journal of Banking and Finance*, **26**,1473–1486.

Furman, E. and Landsman, Z. (2006). Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin*, **36**, 433–462.

Grechuk, B., Molyboha, A. and Zabarankin, M. (2012). Mean-deviation analysis in the theory of choice. *Risk Analysis*, **32**(8), 1277–1292.

Grechuk, B., Molyboha, A. and Zabarankin, M. (2013). Cooperative games with general deviation measures. *Mathematical Finance*, **23**(2), 339–365.

Grechuk, B. and Zabarankin, M. (2012). Optimal risk sharing with general deviation measures. *Annals of Operations Research*, **200**(1), 9–21.

Han, X., Wang, B., Wang, R. and Wu, Q. (2023). Risk concentration and the mean-Expected Shortfall criterion. *Mathematical Finance*, forthcoming.

Herdegen, M. and Khan, N. (2022a). Mean-ρ portfolio selection and ρ-arbitrage for coherent risk measures. *Mathematical Finance*, **32**(1), 226–272.

Herdegen, M. and Khan, N. (2024). ρ-arbitrage and ρ-consistent pricing for star-shaped risk measures. *Mathematics of Operations Research*, forthcoming.

Jones, B. L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal*, **7**(4), 44–54.

Kaina, M. and Rüschendorf, L. (2009). On convex risk measures on L^p-spaces. *Mathematical Methods of Operations Research*, **69**, 475–495.

Krätschmer, V., Schied, A. and Zähle, H. (2014). Comparative and qualitative robustness for law-invariant risk measures. *Finance and Stochastics*, **18**(2), 271–295.

Kuhn, D., Esfahani, P. M., Nguyen, V. A. and Shafieezadeh-Abadeh, S. (2019). Wasserstein distributionally robust optimization: theory and applications in machine learning. *INFORMS TutORials in Operations Research*, 130–166.

Laeven, R. J., Gianin, E. R. and Zullino, M. (2024). Law-invariant return and star-shaped risk measures. *Insurance: Mathematics and Economics*, **117**, 140–153.

Laeven, R. J. and Stadje, M. (2013). Entropy coherent and entropy convex measures of risk. *Mathematics of Operations Research*, **38**(2), 265–293.

Liu, F., Cai, J., Lemieux, C. and Wang, R. (2020). Convex risk functionals: Representation and applications. *Insurance: Mathematics and Economics*, **90**, 66–79.

Maccheroni, F., Marinacci, M., Rustichini, A. and Taboga, M. (2009). Portfolio selection with monotone mean-variance preferences. *Mathematical Finance*, **19**(3), 487–521.

Mao, T. and Wang, R. (2020). Risk aversion in regulatory capital calculation. *SIAM Journal on Financial Mathematics*, **11**(1), 169–200.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, **7**(1), 77–91.

McNeil, A. J., Frey, R. and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools.* Revised Edition. Princeton, NJ: Princeton University Press.

Newey, W. and Powell, J. (1987). Asymmetric least squares estimation and testing. *Econometrica*, **55**(4), 819–847.

Nie, B., Tian, D., and Jiang, L. (2024). Set-valued star-shaped risk measures. *arXiv*: 2402.18014.

Pesenti, S., Wang, Q. and Wang R. (2024). Optimizing distortion riskmetrics with distributional uncertainty.

Mathematical Programming, forthcoming.
Rockafellar, R. T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.
Rockafellar, R. T. and Uryasev, S. (2013). The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, **18**(1–2), 33–53.
Rockafellar, R. T., Uryasev, S. and Zabarankin, M. (2006). Generalized deviations in risk analysis. *Finance and Stochastics*, **10**(1), 51–74.
Ross, S. A. (1981). Some stronger measures of risk aversion in the small and the large with applications. *Econometrica*, **49**, 621–638.
Rothschild, M. and Stiglitz, J. (1970). Increasing risk I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.
Shaked, M. and Shanthikumar, J.G. (2007). *Stochastic Orders*. Springer Series in Statistics.
Shao, Q. M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability*, 2098–2127.
Sharpe, (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, **19**(3), 425–442.
Simaan, Y. (1997). Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model. *Management Science*, **43**(10), 1437–1446.
Tan, K. S., Wei, P., Wei, W., and Zhuang, S. C. (2020). Optimal dynamic reinsurance policies under a generalized Denneberg's absolute deviation principle. *European Journal of Operational Research*, **282**(1), 345–362.
Wang, Q., Wang, R. and Wei, Y. (2020a). Distortion riskmetrics on general spaces. *ASTIN Bulletin*, **50**(3), 827–851.
Wang, R., Bignozzi, V. and Tsakanas, A. (2015). How superadditive can a risk measure be? *SIAM Jounral on Financial Mathematics*, **6**, 776–803.
Wang, R., Wei, Y. and Willmot, G. E. (2020b). Characterization, robustness, and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.
Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**, 1413–1429.
Williamson, R. E. (1956). Multiply monotone functions and their Laplace transforms. *Duke Mathematical Journal*, **23**(2), 189–207.
Wu, Q., Li, J. Y. M. and Mao, T. (2022). On generalization and regularization via Wasserstein distributionally robust optimization. *arXiv*:2212.05716.
Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

A Monotonicity of mean-deviation models

In this appendix, we provide a characterization for [M] of mean-deviation models. Recall the necessary condition in Lemma 1, that is $\lambda D \in \overline{\mathcal{D}}^p$ for some $\lambda > 0$. This condition prompts us to focus on all deviation measures in $\overline{\mathcal{D}}^p$ in the characterization result. For $D \in \overline{\mathcal{D}}^p$, it satisfies the following range-normalized property:

$$\sup_{X \in (L^p)^\circ} \frac{D(X)}{\text{ess-sup}X - \mathbb{E}[X]} = 1. \tag{S.1}$$

Proposition S.1. *Fix $p \in [1, \infty]$. Let $D \in \overline{\mathcal{D}}^p$ satisfy (S.1), and let $U = V(\mathbb{E}, D)$ be defined by (7) with $U(X) < \infty$ for all $X \in L^p$. Suppose that either $V : \mathbb{R} \times [0, \infty)$ is left continuous in its second argument or the maximizer in (S.1) is attainable. Then, U satisfies [M] if and only if $V(m-a, d+a) \leqslant V(m, d)$ for all $m \in \mathbb{R}$ and $a, d \geqslant 0$.*

Proof. While the proof is similar to that of Proposition 4 in Grechuk et al. (2012), which establishes the same necessary and sufficient condition under the assumption that $U = V(\mathbb{E}, D)$ is continuous, we provide the full details here for the sake of completeness and clarity.

We first verify sufficiency. For $X, Y \in L^p$ satisfying $X \leqslant Y$ and $X \neq Y$, we denote by $a = \mathbb{E}[Y] - \mathbb{E}[X] > 0$, and it holds that $\text{ess-sup}(X - Y) \leqslant 0$. By (S.1), we have

$$D(X - Y) \leqslant \text{ess-sup}(X - Y) - \mathbb{E}[X - Y] \leqslant a.$$

Since D satisfies (D4), we have

$$D(X) \leqslant D(Y) + D(X - Y) \leqslant D(Y) + a. \tag{S.2}$$

Therefore,

$$\begin{aligned} U(X) = V(\mathbb{E}[X], D(X)) &= V(\mathbb{E}[Y] - a, D(X)) \\ &\leqslant V(\mathbb{E}[Y], D(X) - a) \leqslant V(\mathbb{E}[Y], D(Y)) = U(Y), \end{aligned}$$

where the first inequality follows from the assumption by letting $m = \mathbb{E}[Y]$, $d = D(X) - a$, and the second inequality is due to $D(X) - a \leqslant D(Y)$ in (S.2). Hence, we conclude that U satisfies [M]. Conversely, we first consider the case that V is left continuous in its second argument. It follows from (S.1) that for any $\varepsilon > 0$, there exists $X_1 \in (L^p)^\circ$ such that

$$1 - \varepsilon \leqslant \frac{D(X_1)}{\text{ess-sup}X_1 - \mathbb{E}[X_1]} \leqslant 1.$$

Let $m \in \mathbb{R}$ and $d, a \geqslant 0$. We define

$$X_2 = a\frac{X_1 - \text{ess-sup}X_1}{\text{ess-sup}X_1 - \mathbb{E}[X_1]} \quad \text{and} \quad X_3 = \frac{d}{a}X_2 + m + d.$$

Through standard calculation, $\mathbb{E}[X_3] = m$, $(1-\varepsilon)d \leqslant D(X_3) \leqslant d$, $\mathbb{E}[X_2 + X_3] = m - a$ and $(a+d)(1-\varepsilon) \leqslant D(X_2 + X_3) \leqslant a + d$. Note that $X_2 \leqslant 0$ which implies $X_2 + X_3 \leqslant X_3$. Using [M], we have

$$V(m - a, (a + d)(1 - \varepsilon)) \leqslant V(\mathbb{E}[X_2 + X_3], D(X_2 + X_3)) \leqslant V(\mathbb{E}[X_3], D(X_3)) \leqslant V(m, d).$$

Letting $\varepsilon \downarrow 0$ and using the left continuity, we conclude that $V(m-a, a+d) \leqslant V(m, d)$ for for all $m \in \mathbb{R}$ and $a, d \geqslant 0$. Now, we assume that the maximizer in (S.1) is attainable, and the necessity follows a similar proof to the previous arguments by constructing X_1 such that $D(X_1)/(\text{ess-sup}X_1 - \mathbb{E}[X_1]) = 1$. Hence, we complete the proof. □

We note that the ES-deviation $\text{ES}_\alpha - \mathbb{E}$ for $\alpha \in (0, 1)$ serves as an example where the maximizer in (S.1) is attainable.

B Axiomatization of monotonic mean-deviation measures

This appendix contains details on the axiomatization of monotonic mean-deviation measures, and its connection to the results of Grechuk et al. (2012). We first present two weaker axioms than A1 and A2, respectively.

B1 If $c_1 \leqslant c_2$, then $c_1 \succeq c_2$ for any $c_1, c_2 \in \mathbb{R}$.

B2 For any X, Y satisfying $\mathbb{E}[X] = \mathbb{E}[Y]$ and $c > 0$, $X \succeq Y$ if and only if $X + c \succeq Y + c$.

Grechuk et al. (2012) established a representation result for mean-deviation models using Axioms A1, B2 and A3–A7. To obtain the characterization in Theorem 2, we first use Axioms B1 and A2–A7 to characterize the preferences that can be represented by the form of MD_g^D in (10), where g and D are not necessarily 1-Lipschitz continuous and weakly upper range-dominated, respectively. Comparing to monotonic mean-deviation measures, this class of mappings further include the mean-variance model that does not satisfy [M].

Proposition S.2. *Fix $p \in [1,\infty]$. A preference $\succeq$ satisfies Axioms B1 and A2–A7 if and only if $\succeq$ can be represented by $\mathrm{MD}_g^D = g \circ D + \mathbb{E}$ where $D \in \mathcal{D}^p$ is continuous and $g : [0,\infty) \to \mathbb{R}$ is continuous and strictly increasing.*

Proof of Proposition S.2. We first show sufficiency. Let $\mathrm{MD}_g^D = g \circ D + \mathbb{E}$ represent $\succeq$ where $D \in \mathcal{D}^p$ and $g : [0,\infty) \to \mathbb{R}$ is some continuous, non-constant and increasing function. Axioms B1, A2, A3, A5 are straightforward by the properties of $D \in \mathcal{D}^p$.

The condition that $X \leqslant_{\mathrm{cx}} Y$ implies $X \succeq Y$ in Axiom A4 comes from Theorem 4.1 of Dana (2005) which showed that every law-invariant continuous convex measure on an atomless probability space is consistent with convex ordering. Moreover, for any $X \in (L^p)^\circ$, since $D(X) > 0$, together with the fact that g is a strictly increasing function, we have $g(D(X)) > g(0)$. Therefore, we have $\mathrm{MD}_g^D(X) > \mathrm{MD}_g^D(\mathbb{E}[X])$, which implies that $\mathbb{E}[X] \succ X$ for any $X \in (L^p)^\circ$. Hence, we have verified Axiom A4.

To show Axiom A6 for MD_g^D, for any $X, Y \in L^p$ such that $\mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathrm{MD}_g^D(X) = \mathrm{MD}_g^D(Y)$, we have $g(D(X)) = g(D(Y))$ and thus $D(X) = D(Y)$ because g is strictly increasing. In this case, for any $\lambda > 0$, we have

$$\begin{aligned}
\mathrm{MD}_g^D(\lambda X + (1-\lambda)Y) &= g(D(\lambda X + (1-\lambda)Y)) + \lambda\mathbb{E}[X] + (1-\lambda)\mathbb{E}[Y] \\
&\leqslant g(\lambda D(X) + (1-\lambda)D(Y)) + \mathbb{E}[Y] \\
&\leqslant g(D(X)) + \mathbb{E}[X] = \mathrm{MD}_g^D(X).
\end{aligned}$$

Axiom A7 follows directly from the fact that D and g are continuous.

Next, we prove necessity. Axioms B1 and A2, A3 and A5 imply the existence of a unique certainty equivalence functional $\rho : L^p \to \mathbb{R}$, i.e., we have $X \succeq Y \iff \rho(X) \leqslant \rho(Y)$ for any $X, Y \in L^p$, and $\rho(c) = c$ for any $c \in \mathbb{R}$; see Theorem 3.3 of Alcantud et al. (2003). In particular, ρ is continuous according to Axiom A7.

Let $X_0 \in (L^p)^\circ$ be such that $\mathbb{E}[X_0] = 0$. Define $\phi(\lambda) = \rho(\lambda X_0)$ for $\lambda \geqslant 0$. We have $\phi(0) = \rho(0)$. The continuity of ϕ follows from the continuity of ρ. Since $\mathbb{E}[\lambda X_0] \succ \lambda X_0$ for any $\lambda > 0$ by Axiom 5, we have $\rho(0) = \rho(\mathbb{E}[\lambda X_0]) < \rho(\lambda X_0)$. This implies that $\phi(\lambda) > \phi(0)$ for any $\lambda > 0$. Moreover, it follows from Axiom 5 that $\rho(\lambda_1 X_0) \leqslant \rho(\lambda_2 X_0)$ for any $0 < \lambda_1 < \lambda_2$ as $\lambda_1 X_0 \leqslant_{\mathrm{cx}} \lambda_2 X_0$. Thus, we have $\phi(\lambda_1) \leqslant \phi(\lambda_2)$ which implies that ϕ is an increasing function on $[0,\infty)$. To show the inequality is strict, we assume by the contradiction, i.e., $\phi(\lambda_1) = \phi(\lambda_2)$. In this case, we have $\rho(\lambda_1 X_0) = \rho(\lambda_2 X_0)$ and $\rho(\lambda_1 k X_0) = \rho(\lambda_2 k X_0)$ for any $k > 0$ by Axiom A3. Let $k = \lambda_1/\lambda_2 < 1$. By induction, we have $\rho(\lambda_1 X_0) = \rho(\lambda_1 k^n X_0)$ for any $n \in \mathbb{N}$. Letting $n \to \infty$, by Axiom A7, we have $\rho(\lambda_1 X_0) = \rho(0)$, which contradicts to Axiom A4. Thus, ϕ is a strictly increasing and continuous function on $[0,\infty)$, and its inverse function $\phi^{-1}(x) := \inf\{\lambda \in [0,\infty) : \phi(\lambda) \geqslant x\}$ is also strictly increasing and continuous on the range of ϕ.

For $X \in L^p$, let $\overline{X} = X - \mathbb{E}[X]$ and $D(X) = \phi^{-1}(\rho(\overline{X}))$. Since ρ and ϕ are continuous functions, we know that D is continuous. Next, we aim to verify that $D \in \mathcal{D}^p$. It is clear that D is law-invariant since ρ is law-invariant by Axiom A4, and thus (D5) holds. For any $c \in \mathbb{R}$, $D(X+c) = \phi^{-1}(\rho(\overline{X+c})) = D(X)$, which implies (D1). Note that Axiom A4 implies $\rho(\overline{X}) > \rho(0)$ for all $X \in (L^p)^\circ$. We have $D(X) = \phi^{-1}(\rho(\overline{X})) > \phi^{-1}(\rho(0)) = 0$ as ϕ^{-1} is strictly increasing. For any $c \in \mathbb{R}$,

$D(c) = \phi^{-1}(\rho(0)) = 0$. Thus, (D2) holds. For any $X \in L^p$, we have

$$\rho(D(X)X_0) = \phi(D(X)) = \phi \circ \phi^{-1}(\rho(\overline{X})) = \rho(\overline{X}) \Longleftrightarrow \overline{X} \simeq D(X)X_0. \tag{S.3}$$

It then follows from Axiom A3 that $\lambda\overline{X} \simeq \lambda D(X)X_0$ for all $\lambda \geqslant 0$. Hence, we have $\rho(\overline{\lambda X}) = \rho(\lambda D(X)X_0)$. On the other hand, $\overline{\lambda X} \simeq D(\lambda X)X_0$ implies $\rho(\overline{\lambda X}) = \rho(D(\lambda X)X_0)$. This concludes that $\rho(\lambda D(X)X_0) = \rho(D(\lambda X)X_0)$, which is equivalent to $\phi(\lambda D(X)) = \phi(D(\lambda X))$. Note that ϕ is strictly increasing. It holds that $\lambda D(X) = D(\lambda X)$ which implies (D3). For $X, Y \in L^p$, if X or Y is constant, (D4) holds directly. Otherwise, we have $D(X) > 0$ and $D(Y) > 0$. Combining (S.3) and Axiom A3, we have $\rho(\overline{X}/D(X)) = \rho(X_0)$ and $\rho(\overline{Y}/D(Y)) = \rho(X_0)$ which implies $\rho(\overline{X}/D(X)) = \rho(\overline{Y}/D(Y))$. Moreover, by Axiom A6, for all $\lambda \in [0, 1]$,

$$\rho\left(\lambda\frac{\overline{X}}{D(X)} + (1-\lambda)\frac{\overline{Y}}{D(Y)}\right) \leqslant \rho\left(\frac{\overline{Y}}{D(Y)}\right) = \rho(X_0).$$

By setting $\lambda = D(X)/(D(X)+D(Y))$, we have $\rho\left((\overline{X}+\overline{Y})/(D(X)+D(Y))\right) \leqslant \rho(X_0)$. Applying (S.3) and Axiom A3 again, we have the following relation:

$$\frac{\overline{X}+\overline{Y}}{D(X)+D(Y)} = \frac{\overline{X+Y}}{D(X)+D(Y)} \simeq \frac{D(X+Y)X_0}{D(X)+D(Y)}.$$

Hence, denote by $k = D(X+Y)/(D(X)+D(Y))$, and we have $\rho(kX_0) \leqslant \rho(X_0)$, which implies $\phi(k) \leqslant \phi(1)$. Noting that ϕ is strictly increasing, we have $D(X+Y) \leqslant D(X) + D(Y)$ and (D4) holds.

For any $X \in L^p$, using $X \simeq \rho(X)$, we have $X - \mathbb{E}[X] \simeq \rho(X) - \mathbb{E}[X]$ by Axiom A2, which implies $\rho(X - \mathbb{E}[X]) = \rho(X) - \mathbb{E}[X]$. Therefore, using (S.3),

$$\rho(X) = \rho(X - \mathbb{E}[X]) + \mathbb{E}[X] = \rho(\overline{X}) + \mathbb{E}[X] = \phi(D(X)) + \mathbb{E}[X], \quad \text{for all } X \in L^p,$$

where the last step follows from (S.3). This completes the proof. □

Now, we are ready to prove Theorem 2.

Proof of Theorem 2. Sufficiency is straightforward by combining Theorem 1, Lemma 1 and Proposition S.2. Next, we show the necessity. By Proposition S.2, $\succeq$ can be represented by $\mathrm{MD}_f^{D'} = f \circ D' + \mathbb{E}$ where $D' \in \mathcal{D}^p$, and $f : [0, \infty) \to \mathbb{R}$ is some continuous and strictly increasing function. Since $\mathrm{MD}_f^{D'}$ satisfies monotonicity, by Lemma 1, we have $D' \in \overline{\mathcal{D}}_K^p$. Define $g = f \circ D$ and $D = D'/K$, we have $\mathrm{MD}_f^{D'} = \mathrm{MD}_g^{D} = g \circ D + \mathbb{E}$ where $g : [0, \infty) \to \mathbb{R}$ is some continuous and strictly increasing function and $D \in \overline{\mathcal{D}}^p$. By Theorem 1, g is 1-Lipschitz continuous. Hence, we complete the proof. □

C Proof of Theorem 5

This appendix contains the detailed proof of Theorem 5.

Proof. The Law of Large Numbers yields $\widehat{x}_n \xrightarrow{\mathbb{P}} \mathbb{E}[X]$. By Theorem 2.6 of Krätschmer et al. (2014), the empirical estimator for a finite convex risk measure on L^p is consistent, that is, $\widehat{D}(n) + \widehat{x}_n \xrightarrow{\mathbb{P}} D(X) + \mathbb{E}[X]$, and this gives $\widehat{D}(n) \xrightarrow{\mathbb{P}} D(X)$. Moreover, since $g \in \mathcal{G}$, we have $g(\widehat{D}(n)) + \widehat{\mathbb{E}}[n] \xrightarrow{\mathbb{P}} g(D(X)) + \mathbb{E}[X]$. This proves the first part of the result.

Next, we will show the asymptotic normality. Let $B = (B_t)_{t\in[0,1]}$ be a standard Brownian bridge, and let $d_n = \sqrt{n}(\widehat{D}(n) - D(X))$, $e_n = \sqrt{n}(\widehat{x}_n - \mathbb{E}[X])$, and $g_n = \sqrt{n}(g(\widehat{D}(n)) - g(D(X)))$. We need to first show

$$(d_n, e_n) \xrightarrow{\mathrm{d}} (Z, W) := \left(\int_0^1 \frac{B_s h'(1-s)}{\tilde{f}(s)} \mathrm{d}s, \int_0^1 \frac{B_s}{\tilde{f}(s)} \mathrm{d}s\right). \tag{S.4}$$

By the Cramér-Wold theorem, it is sufficient to show

$$a d_n + b e_n \xrightarrow{\mathrm{d}} aZ + bW \text{ for all } a, b \in \mathbb{R}. \tag{S.5}$$

Note that $aD + b\mathbb{E}$ can be written as an integral of the quantile, that is,

$$aD(X) + b\mathbb{E}[X] = \int_0^1 F^{-1}(t)(ah'(1-t) + b)\mathrm{d}t.$$

Denote by A_n the empirical quantile process, that is,

$$A_n(t) = \sqrt{n}(\widehat{F}_n^{-1}(t) - F^{-1}(t)), \quad t \in (0,1).$$

It follows that

$$a d_n + b e_n = \int_0^1 A_n(t)(ah'(1-t) + b)\mathrm{d}t.$$

Using this representation, the convergence (S.5) can be verified via Theorem 3.2 of Jones and Zitikis (2003), and we briefly verify it here. It is well known that, under Assumption 1, as $n \to \infty$, A_n converges to the Gaussian process $B/\tilde{f}$ in $L^\infty[1-\delta, 1+\delta]$ for any $\delta > 0$ (see e.g., Del Barrio et al. (2005)). This yields

$$\int_\delta^{1-\delta} A_n(t)(ah'(1-t) + b)\mathrm{d}t \xrightarrow{\mathrm{d}} \int_\delta^{1-\delta} \frac{B_t}{\tilde{f}(t)}(ah'(1-t) + b)\mathrm{d}t.$$

To show (S.5), it suffices to verify

$$\int_\delta^{1-\delta} \frac{B_t}{\tilde{f}(t)}(ah'(1-t) + b)\mathrm{d}t \to \int_0^1 \frac{B_t}{\tilde{f}(t)}(ah'(1-t) + b)\mathrm{d}t \quad \text{as } \delta \downarrow 0. \tag{S.6}$$

Denote by $w_t = t(1-t)$. Since $h \in \Phi^p$ and $X \in L^\gamma$, we have, for some $C > 0$,

$$|h'(1-t)| \leqslant C w_t^{1/p-1}; \quad |F^{-1}(t)| \leqslant C w_t^{-1/\gamma}; \quad \frac{1}{\tilde{f}(t)} = \frac{\mathrm{d}F^{-1}(t)}{\mathrm{d}t} \leqslant C w_t^{-1/\gamma-1}.$$

Note that $1/p - 1/\gamma > 1/2$ and $B_t = o_{\mathbb{P}}(w_t^{1/2-\varepsilon})$ for any $\varepsilon > 0$ as $t \to 0$ or 1. Hence, for some $\eta > 0$,

$$\left| B_t \frac{ah'(1-t) + b}{\tilde{f}(t)} \right| = o_{\mathbb{P}}(w_t^{\eta-1}) \quad \text{for } t \in (0,1),$$

and this guarantees (S.6). Therefore, (S.4) holds.

By the Mean Value Theorem, there exists x_0 between $D(X)$ and $\widehat{D}(n)$ such that

$$\sqrt{n}(g(\widehat{D}(n)) - g(D(X))) = g'(x_0)\sqrt{n}(\widehat{D}(n) - D(X)).$$

Using the fact that $\widehat{D}(n) \xrightarrow{\mathbb{P}} D(X)$, we get

$$(g_n, e_n) \xrightarrow{\mathrm{d}} \left(g'(D(X)) \int_0^1 \frac{B_s h'(1-s)}{\tilde{f}(s)} \mathrm{d}s, \int_0^1 \frac{B_s}{\tilde{f}(s)} \mathrm{d}s \right).$$

Hence,

$$\sqrt{n} \left(\widehat{\mathrm{MD}}_g^D(n) - \mathrm{MD}_g^D(X) \right) = g_n + e_n \xrightarrow{\mathrm{d}} g'(D(X)) \int_0^1 \frac{B_s h'(1-s)}{\tilde{f}(s)} \mathrm{d}s + \int_0^1 \frac{B_s}{\tilde{f}(s)} \mathrm{d}s,$$

or equivalently,

$$\sqrt{n} \left(\widehat{\mathrm{MD}}_g^D(n) - \mathrm{MD}_g^D(X) \right) \xrightarrow{\mathrm{d}} \int_0^1 \frac{B_s}{\tilde{f}(s)} (h'(1-s) g'(D(X)) + 1) \mathrm{d}s.$$

Using the convariance property of the Brownian bridge, that is, $\mathrm{Cov}(B_t, B_s) = s - st$ for $s < t$, we have

$$\begin{aligned}
&\mathrm{Var}\left[\int_0^1 \frac{B_s(h'(1-s)g'(D(X)) + 1)}{\tilde{f}(s)} \mathrm{d}s \right] \\
&= \mathbb{E}\left[\int_0^1 \int_0^1 \frac{(h'(1-s)g'(D(X)) + 1)(h'(1-t)g'(D(X)) + 1) B_s B_t}{\tilde{f}(s)\tilde{f}(t)} \mathrm{d}t \mathrm{d}s \right] \\
&= \int_0^1 \int_0^1 \frac{(h'(1-s)g'(D(X)) + 1)(h'(1-t)g'(D(X)) + 1)(s \wedge t - st)}{\tilde{f}(s)\tilde{f}(t)} \mathrm{d}t \mathrm{d}s.
\end{aligned}$$

Thus, $\sqrt{n} \left(\widehat{\mathrm{MD}}_g^D(n) - \mathrm{MD}_g^D(X) \right) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_g^2)$ in which σ_g^2 is given by (18). □

D Worst-case values under model uncertainty

In the context of robust risk evaluation, one may only have partial information on a risk X to be evaluated. In this section, we discuss two model uncertainty problems—mean-variance uncertainty and Wasserstein uncertainty—using MD_g^D as the criterion.

D.1 Mean-variance uncertainty

First, consider the case in which one only knows the mean and the variance of X. This setup has wide applications in model uncertainty and portfolio optimization. Denote by

$$L^2(m, v) = \left\{ X \in L^2 : \mathbb{E}[X] = m,\ \sigma^2(X) = v^2 \right\}.$$

For a fixed $g \in \mathcal{G}$ and $D \in \mathcal{D}^p$ with $p \in [1, 2]$, we consider the following worst-case problem

$$\overline{\mathrm{MD}}_g^D(m, v) = \sup \left\{ \mathrm{MD}_g^D(X) : X \in L^2(m, v) \right\}. \tag{S.7}$$

Since g is increasing, the method for solving (S.7) is similar to those employed for related worst-case risk measures studied in the literature; see, for instance, Liu et al. (2020), Pesenti et al. (2024), and Bernard et al. (2024). In particular, when $g(x) = \lambda x$ for some $0 < \lambda \leqslant 1$, (S.7) simplifies to the problem analyzed in Section 5 of Pesenti et al. (2024). For completeness, we provide a detailed analysis of our framework.

We collect some notations in Section 6 that are needed here. Define

$$\mathcal{H} = \{h: \ h \text{ maps from } [0,1] \text{ to } \mathbb{R} \text{ is of bounded variation with } h(0)=h(1)=0\};$$
$$D_h(X) = \int_0^1 \mathrm{VaR}_\alpha(X) h'(1-\alpha)\mathrm{d}\alpha = \int_{\mathbb{R}} h\left(\mathbb{P}(X>x)\right)\mathrm{d}x, \quad h\in\mathcal{H};$$
$$\Phi^p = \{h\in\mathcal{H}: \ h \text{ is concave and } \|h'\|_q<\infty\}, \quad p\in[1,\infty),$$

where $\|h'\|_q = \int_0^1 |h'(s)|\mathrm{d}s$ and $q=(1-1/p)^{-1}$. By Theorem 2.4 of Liu et al. (2020), for $D\in\mathcal{D}^p$ and $p\in[1,\infty)$, there exists a subset of Φ^p, denoted as Ψ^p, such that

$$D(X) = \sup_{h\in\Psi^p}\left\{\int_0^1 \mathrm{VaR}_\alpha(X)h'(1-\alpha)\mathrm{d}\alpha\right\} = \sup_{h\in\Psi^p} D_h(X), \quad X\in L^p, \tag{S.8}$$

Proposition S.3. *Given $p\in[1,2]$, $m\in\mathbb{R}$, $v>0$, $g\in\mathcal{G}$ and $D\in\mathcal{D}^p$ defined by (S.8), we have*

$$\overline{\mathrm{MD}}_g^D(m,v) = \sup_{h\in\Psi^p} g\left(v\left\|h'\right\|_2\right)+m.$$

Proof. By (S.8), we have

$$\begin{aligned}\overline{\mathrm{MD}}_g^D(m,v) &= \sup_{X\in L^2(m,v)} g(D(X))+m\\ &= \sup_{X\in L^2(m,v)} g\left(\sup_{h\in\Psi^p}\left\{\int_0^1 \mathrm{VaR}_\alpha(X)h'(1-\alpha)\mathrm{d}\alpha\right\}\right)+m\\ &= \sup_{X\in L^2(m,v)} g\left(\sup_{h\in\Psi^p} D_h(X)\right)+m = g\left(\sup_{h\in\Psi^p}\sup_{X\in L^2(m,v)} D_h(X)\right)+m,\end{aligned}$$

where the last step holds because g is increasing. By Theorem 3.1 of Liu et al. (2020), we have that $\sup_{X\in L^2(m,v)} D_h(X) = v\left\|h'\right\|_2$ for any $h\in\Phi^p$. This completes the proof. □

Remark S.1. The worst-case problem formulated in (S.7) can be extended to the case of other central moment instead of the variance. For $a>1, m\in\mathbb{R}$ and $v>0$, denote by

$$L^a(m,v) = \left\{X\in L^a : \mathbb{E}[X]=m,\ \mathbb{E}\left[|X-m|^a\right]=v^a\right\}.$$

Suppose that $p\in[1,a]$. Theorem 5 of Pesenti et al. (2024) implies that

$$\sup\left\{D_h(X) : X\in L^p(m,v)\right\} = v[h]_q, \quad h\in\Phi_p.$$

Therefore, for $D\in\mathcal{D}^p$ defined by (S.8), it follows the similar arguments in the proof of Proposition S.3 that

$$\left\{\mathrm{MD}_g^D(X) : X\in L^a(m,v)\right\} = \sup_{h\in\Psi^p} g\left(v[h]_q\right)+m.$$

Example S.1. Let $D=\mathrm{ES}_\alpha-\mathbb{E}$ with $\alpha\in(0,1)$. We have $D=D_h$, where $h(t)=(t-\alpha)_+/(1-\alpha)-t$ for $t\in[0,1]$. It holds that

$$[h]_q = \min_{x\in\mathbb{R}}\|h'-x\|_q = \min_{x\in\mathbb{R}}\left(\alpha|1+x|^q+(1-\alpha)\left|\frac{\alpha}{1-\alpha}-x\right|^q\right)^{1/q}.$$

By standard manipulation, we conclude that the minimizer of the above optimization problem can be attained at $x^* = (\alpha(1-\alpha)^{p-2} - \alpha^{p-1})/(\alpha^{p-1} + (1-\alpha)^{p-1})$, and the optimal value is $[h]_q = \alpha\left(\alpha^p(1-\alpha) + \alpha(1-\alpha)^p\right)^{-1/p}$. Thus, in this case, we have

$$\overline{\mathrm{MD}}_g^D(m,v) = m + g\left(v\alpha\left(\alpha^p(1-\alpha) + \alpha(1-\alpha)^p\right)^{-1/p}\right).$$

We compare the results for the normal, Pareto and exponential distributions with the worst-case distribution with the same mean and variance. Setting $p=2$ and both the mean and variance to 1, we show the values of MD_g^D and $\overline{\mathrm{MD}}_g^D$ when $D = \mathrm{ES}_\alpha - \mathbb{E}$ for different values of $\alpha \in [0.9, 0.99]$ in Figure S.1. In particular, when $g(x) = x$, MD_g^D simplifies to ES_α. Given that g is 1-Lipschitz continuous, it is expected that the worst-case values of ES_α will be larger than those of $\overline{\mathrm{MD}}_g^D$ for other forms of g.

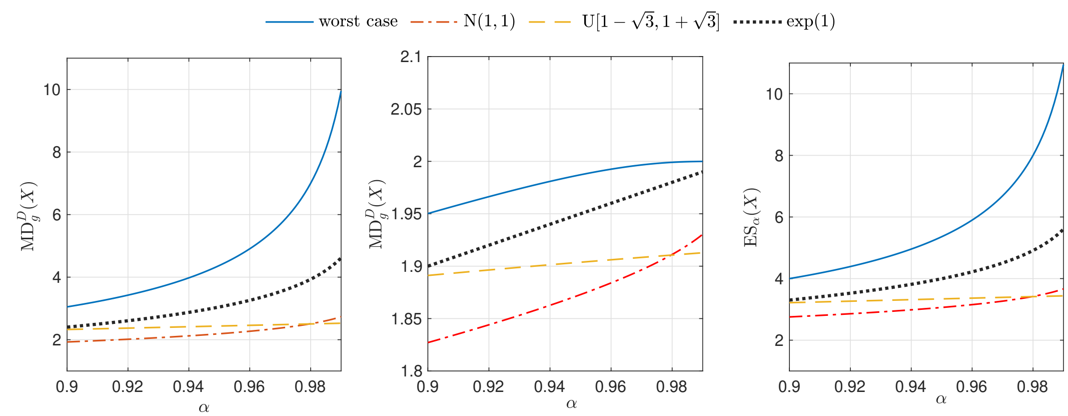


Figure S.1: Values of MD_g^D and $\overline{\mathrm{MD}}_g^D$ with $g(x) = x + e^{-x} - 1$ (left), $g(x) = 1 - e^{-x}$ (middle) and $g(x) = x$ (right)

D.2 Wasserstein uncertainty

Optimization problems under the uncertainty set of a Wasserstein ball are also common in the literature when quantifying the discrepancy between a benchmark distribution and alternative scenarios; see e.g., Esfahani and Kuhn (2018). For two distributions F and G, the type-p Wasserstein metric with $p \geqslant 1$, is given by

$$W_p(F,G) = \left(\int_0^1 \left|F^{-1}(u) - G^{-1}(u)\right|^p \,\mathrm{d}u\right)^{1/p}.$$

Denote by $\mathcal{M}_p$ the set of all distribution functions that have finite pth moment. For $F_0 \in \mathcal{M}_p$ and $\varepsilon \geqslant 0$, we define the following uncertainty set based on the type-p Wasserstein metric

$$\mathcal{B}_p(F_0, \varepsilon) = \{F \in \mathcal{M}_p : W_p(F, F_0) \leqslant \varepsilon\}.$$

The above uncertainty set is also known as a type-p Wasserstein ball (see e.g., Kuhn et al. (2019), Wu et al. (2022) and Bernard et al. (2024)), where F_0 is the center and ε is the radius. Note that $\varepsilon = 0$ corresponds to the case of no model uncertainty. In what follows, we focus on the type-2 Wasserstein ball. For any $\varepsilon \geqslant 0$, $g \in \mathcal{G}$ and $D \in \mathcal{D}^p$ with $p \in [1,2]$, we define the worst-case MD_g^D under type-2 Wasserstein ball as

$$\widetilde{\mathrm{MD}}_g^D(X|\varepsilon) = \sup\left\{\mathrm{MD}_g^D(Y) : F_Y \in \mathcal{B}_2(F_X, \varepsilon)\right\}.$$

The following result gives a formula to compute the above worst-case problems.

Proposition S.4. *Given $p \in [1,2]$, $g \in \mathcal{G}$ and $D = D_h$ with $h \in \Phi^p$, we have*

$$\widetilde{\mathrm{MD}}_g^D(X|\varepsilon) = \sup_{t\in[-1,1]} \left\{ g\left(\varepsilon\sqrt{1-t^2}\, \|h'\|_2 + D(X) \right) + t\varepsilon + \mathbb{E}[X] \right\}, \quad \varepsilon \geqslant 0,\ X \in L^2.$$

Proof. Denote by $\mathcal{M} = \{F_Y : Y \in L^2,\ \|Y-X\|_2 \leqslant \varepsilon\}$. We first aim to show that $\mathcal{M} = \mathcal{B}_2(F_X, \varepsilon)$. Note that $\mathcal{B}_2(F_X,\varepsilon) = \{F_Y : Y \in L^2,\ \int_0^1 |F_Y^{-1}(u) - F_X^{-1}(u)|^2 \mathrm{d}u \leqslant \varepsilon^2\}$. It is obvious that $\mathcal{M} \subseteq \mathcal{B}_2(F_X,\varepsilon)$ since $\|Y-X\|_2^2 \geqslant \int_0^1 |F_Y^{-1}(u) - F_X^{-1}(u)|^2 \mathrm{d}u$ for any $X, Y \in L^2$. To see the converse direction, for any $F \in \mathcal{B}_2(F_X,\varepsilon)$, let $Y \in L^2$ be such that Y and X are comonotonic and Y has distribution F. It holds that $\|Y-X\|_2^2 = \int_0^1 |F^{-1}(u) - F_X^{-1}(u)|^2 \mathrm{d}u \leqslant \varepsilon^2$, where the last step is due to $F \in \mathcal{B}_2(F_X,\varepsilon)$. Hence, we have $F \in \mathcal{M}$. This implies that $\mathcal{B}_2(F_X,\varepsilon) \subseteq \mathcal{M}$, and we have concluded that $\mathcal{M} = \mathcal{B}_2(F_X,\varepsilon)$. Note that MD_g^D is law-invariant. We have

$$\begin{aligned} \widetilde{\mathrm{MD}}_g^D(X|\varepsilon) &= \sup\{\mathrm{MD}_g^D(Y) : F_Y \in \mathcal{B}_2(F_X,\varepsilon)\} \\ &= \sup_{\|Y-X\|_2 \leqslant \varepsilon} \mathrm{MD}_g^D(Y) = \sup_{\|Y-X\|_2\leqslant\varepsilon} \{g(D(Y)) + \mathbb{E}[Y]\}. \end{aligned} \tag{S.9}$$

Denote by $\mu_0 = \mathbb{E}[X]$. It holds that

$$\{\mathbb{E}[Y] : \|Y-X\|_2 \leqslant \varepsilon\} = \{\mu_0 + \mathbb{E}[V] : \|V\|_2 \leqslant \varepsilon\} \subseteq [\mu_0 - \varepsilon, \mu_0 + \varepsilon].$$

Therefore, (S.9) reduces to

$$\begin{aligned} \sup_{\|Y-X\|_2\leqslant\varepsilon} \{g(D(Y)) + \mathbb{E}[Y]\} &= \sup_{\mu\in[\mu_0-\varepsilon,\mu_0+\varepsilon]} \sup_{\|Y-X\|_2\leqslant\varepsilon,\ \mathbb{E}[Y]=\mu} \{g(D(Y)) + \mathbb{E}[Y]\} \\ &= \sup_{\mu\in[\mu_0-\varepsilon,\mu_0+\varepsilon]} \sup_{\|V\|_2\leqslant\varepsilon,\ \mathbb{E}[V]=\mu-\mu_0} \{g(D(V+X)) + \mu\} \\ &= \sup_{\mu\in[\mu_0-\varepsilon,\mu_0+\varepsilon]} \sup_{\|V\|_2\leqslant\varepsilon,\ \mathbb{E}[V]=\mu-\mu_0} \{g(D(V) + D(X)) + \mu\} \\ &= \sup_{\mu\in[\mu_0-\varepsilon,\mu_0+\varepsilon]} \sup_{\sigma^2(V)\leqslant\varepsilon^2-(\mu-\mu_0)^2,\ \mathbb{E}[V]=\mu-\mu_0} \{g(D(V) + D(X)) + \mu\}, \end{aligned}$$

where the third equality holds because g is increasing and D is subadditive and comonotonic additive, and we can construct V and X to be comonotonic. Since g is increasing, the inner optimization problem is equivalent to maximizing $D(V)$ over $\{V : \sigma^2(V) \leqslant \varepsilon^2 - (\mu-\mu_0)^2,\ \mathbb{E}[V] = \mu - \mu_0\}$. Using the arguments in the proof of Proposition S.3, we have

$$\sup\{D(V) : \sigma^2(V) \leqslant \varepsilon^2 - (\mu-\mu_0)^2,\ \mathbb{E}[V] = \mu - \mu_0\} = \sqrt{\varepsilon^2 - (\mu-\mu_0)^2}\, \|h'\|_2.$$

Therefore, we have

$$\begin{aligned} \widetilde{\mathrm{MD}}_g^D(X|\varepsilon) &= \sup_{\mu\in[\mu_0-\varepsilon,\mu_0+\varepsilon]} \left\{ g\left(\sqrt{\varepsilon^2 - (\mu-\mu_0)^2}\|h'\|_2 + D(X) \right) + \mu \right\} \\ &= \sup_{t\in[-\varepsilon,\varepsilon]} \left\{ g\left(\sqrt{\varepsilon^2 - t^2}\, \|h'\|_2 + D(X) \right) + t + \mu_0 \right\} \\ &= \sup_{t\in[-1,1]} \left\{ g\left(\varepsilon\sqrt{1-t^2}\, \|h'\|_2 + D(X) \right) + t\varepsilon + \mathbb{E}[X] \right\}, \end{aligned}$$

which completes the proof. □

In Proposition S.4, our analysis is confined to the case of type-2 Wasserstein ball and the signed Choquet integral D_h. Working with a general deviation measure D is not more difficult, as it involves only another supremum over Ψ^p by using (S.8). For the general type-p Wasserstein ball with $p \neq 2$, following similar arguments to those used in the proof of Proposition S.4 leads us to

$$\sup\left\{\mathrm{MD}_g^D(Y) : F_Y \in \mathcal{B}_p(F_X, \varepsilon)\right\} = \sup_{\mu \in [\mathbb{E}[X]-\varepsilon, \mathbb{E}[X]+\varepsilon]} \sup_{\substack{V: \|V\|_p \leqslant \varepsilon \\ \mathbb{E}[V] = \mu - \mathbb{E}[X]}} \{g(D(V+X)) + \mu\}. \tag{S.10}$$

We do not have an explicit formula to solve the inner supremum problem on the right-hand side of (S.10). This is because $\|V\|_p$ and $\mathbb{E}[V]$ do not align very well unless $p = 2$.

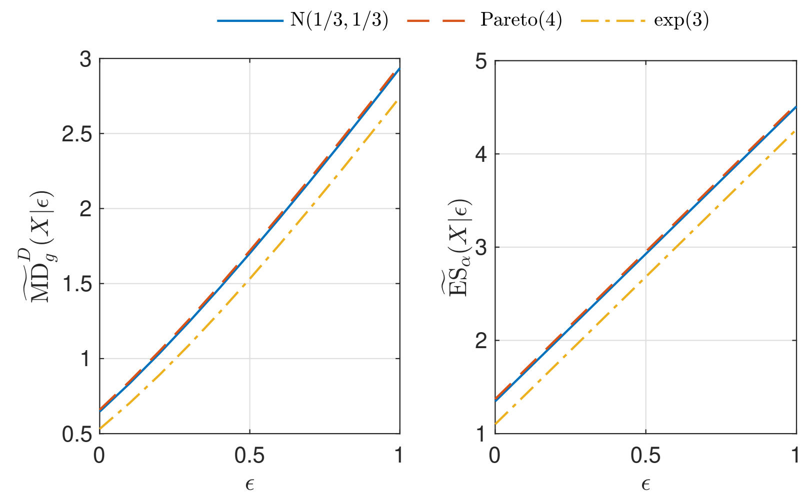


Figure S.2: The values of $\widetilde{\mathrm{MD}}_g^D(X|\varepsilon)$ with $g(x) = x - \log(1+x)$ (left) and $g(x) = x$ (right), and $D = \mathrm{ES}_\alpha - \mathbb{E}$

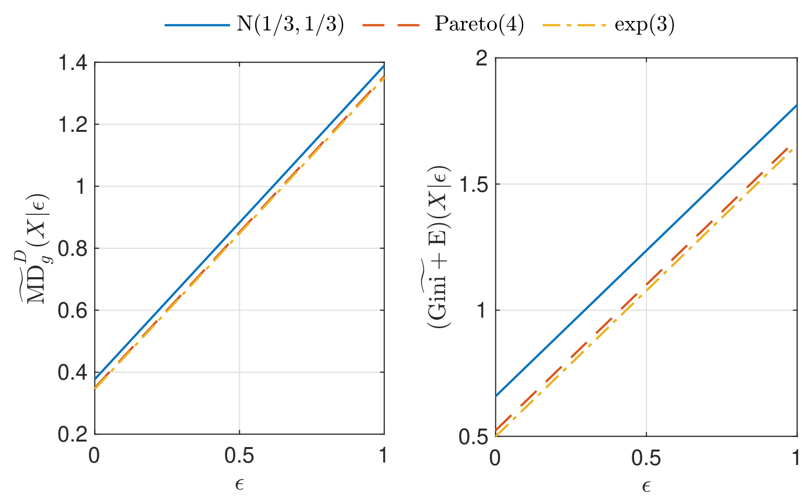


Figure S.3: The values of $\widetilde{\mathrm{MD}}_g^D(X|\varepsilon)$ with $g(x) = x - \log(1+x)$ (left) and $g(x) = x$ (right), and $D = \mathrm{Gini}$

Example S.2. Let $g(x) = x - \log(1+x)$ for $x \in \mathbb{R}$ and $h(t) = (t-\alpha)_+/(1-\alpha) - t$ for $t \in [0,1]$ with

$\alpha \in (0,1)$. We have $D := D_h = \mathrm{ES}_\alpha - \mathbb{E}$, and it follows from Proposition S.4 that

$$\widetilde{\mathrm{MD}}_g^D(X|\varepsilon) = \sup_{t\in[-1,1]} \left\{ g\left(\varepsilon\sqrt{1-t^2}\sqrt{\frac{\alpha}{1-\alpha}} + \mathrm{ES}_\alpha(X) - \mathbb{E}[X] \right) + t\varepsilon + \mathbb{E}[X] \right\}$$

$$= \sup_{t\in[-1,1]} \left\{ t\varepsilon + \varepsilon\sqrt{1-t^2}\sqrt{\frac{\alpha}{1-\alpha}} + \mathrm{ES}_\alpha(X) - \log\left(1 + \varepsilon\sqrt{1-t^2}\sqrt{\frac{\alpha}{1-\alpha}} + \mathrm{ES}_\alpha(X) - \mathbb{E}[X] \right) \right\}.$$

Next, let $D = \mathrm{Gini}$ defined in (19). By Proposition S.4, we have

$$\widetilde{\mathrm{MD}}_g^D(X|\varepsilon) = \sup_{t\in[-1,1]} \left\{ g\left(\frac{\sqrt{3}\varepsilon}{3}\sqrt{1-t^2} + \mathrm{Gini}(X) \right) + t\varepsilon + \mathbb{E}[X] \right\}$$

$$= \sup_{t\in[-1,1]} \left\{ t\varepsilon + \mathbb{E}[X] + \frac{\sqrt{3}\varepsilon}{3}\sqrt{1-t^2} + \mathrm{Gini}(X) - \log\left(1 + \frac{\sqrt{3}\varepsilon}{3}\sqrt{1-t^2} + \mathrm{Gini}(X) \right) \right\}.$$

The maximum values are computed numerically by considering $g(x) = x - \log(1+x)$ and $g(x) = x$ with the benchmark distributions being normal, Pareto and exponential. We calculate the worst values of $\mathrm{MD}_g^D(X)$ when $D = \mathrm{ES}_\alpha - \mathbb{E}$ with $\alpha = 0.9$ and $D = \mathrm{Gini}$, across different values of uncertainty level ε. The results are presented in Figures S.2 and S.3, respectively. Again, when $g(x) = x$, MD_g^D simplifies to ES_α or $\mathrm{Gini} + \mathbb{E}$. Given that g is 1-Lipschitz continuous, the worst-case values of ES_α or $\mathrm{Gini} + \mathbb{E}$ will be larger than those of $\widetilde{\mathrm{MD}}_g^D$ for other forms of g.