Model Aggregation for Risk Evaluation and Robust Optimization

Tiantian Mao* Ruodu Wang† Qinyu Wu‡

Abstract

We introduce a new approach for prudent risk evaluation based on stochastic dominance, which will be called the model aggregation (MA) approach. In contrast to the classic worst-case risk (WR) approach, the MA approach produces not only a robust value of risk evaluation but also a robust distributional model, independent of any specific risk measure. The MA risk evaluation can be computed through explicit formulas in the lattice theory of stochastic dominance, and under some standard assumptions, the MA robust optimization admits a convex-program reformulation. The MA approach for Wasserstein and mean-variance uncertainty sets admits explicit formulas for the obtained robust models. Via an equivalence property between the MA and the WR approaches, new axiomatic characterizations are obtained for the Value-at-Risk (VaR) and the Expected Shortfall (ES, also known as CVaR). The new approach is illustrated with various risk measures and examples from portfolio optimization.

Keywords: Value-at-Risk, Expected Shortfall, stochastic dominance, model aggregation, worst-case risk measures, model uncertainty, robust optimization

1 Introduction

Modern risk management often requires the evaluation of risks under multiple probability measures, called scenarios. The risk evaluation obtained under various scenarios needs to be aggregated properly, and a prudent approach is often implemented in practice. As a prominent example, in the Fundamental Review of the Trading Book (FRTB) of Basel IV (BCBS (2019)), banks need to evaluate the market risk of their portfolio losses under stressed scenarios, in particular including a model generated from data during the financial crisis of 2007, and the obtained risk models are aggregated via a worst-case approach; see Wang and Ziegel (2021, Section 1) for a description of

*Department of Statistics and Finance, University of Science and Technology of China, China. E-mail: tmao@ustc.edu.cn

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

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

the stressed scenarios and the model aggregation in the FRTB, and Cambou and Filipović (2017) for techniques to generate regulatory scenarios. In the literature of portfolio risk assessment and optimization, the worst-case approach is popular; we refer to El Ghaoui et al. (2003), Natarajan et al. (2008), Zhu and Fukushima (2009) and Glasserman and Xu (2014) for robust portfolio optimization, and Embrechts et al. (2013) and Wang et al. (2013) for robust risk aggregation. This idea further leads to many studies in distributionally robust optimization; see Delage and Ye (2010), Gao and Kleywegt (2022), Esfahani and Kuhn (2018) and Blanchet and Murthy (2019). In this paper, we will work in the context where a prudent risk evaluation under multiple models, which is our main focus, is desirable.[1]

A natural question for risk management in the presence of model uncertainty is how to generate a robust model from a collection of models resulting from statistical and machine learning procedures, operational considerations, or expert's opinion. Such a robust model can be used for risk evaluation, simulation, optimization, and decision making.

Our main ideas to address this question are described below. Let $\mathcal{M}$ be a set of distributions on $\mathbb{R}$, representing possible risk models, called distributional models; for illustrative purposes, we focus on one-dimensional financial losses for which the theory of risk measures is rich. Suppose that a risk analyst evaluates a random loss using different methodologies, scenarios or data sets, and obtains a collection $\mathcal{F} \subseteq \mathcal{M}$ of distributional models. Here, the number of models in $\mathcal{F}$ may be finite or infinite. For instance, $\mathcal{F}$ may contain distributions of the random loss under different probability measures (economic scenarios), estimation methods, or values of statistical parameters; alternatively, $\mathcal{F}$ may represent distributions from losses which may occur from different possible decisions from a business competitor. The set $\mathcal{F}$ will be called an *uncertainty set*. The distributions in $\mathcal{F}$ will be used to assess the risk, together with a risk measure $\rho : \mathcal{M} \to \mathbb{R}$, such as a Value-at-Risk (VaR) or an Expected Shortfall (ES, also known as CVaR); see Section 2.2 for their definitions. Prudent regulation and risk management require a conservative approach which aggregates the above information. There are two conceptually intuitive ways to generate a robust assessment of the risk:

(i) Directly calculate the maximum (or supremum) of $\rho(F)$ over $F \in \mathcal{F}$.

(ii) Calibrate a robust (conservative) distributional model F^* based on $\mathcal{F}$, and calculate $\rho(F^*)$.

Arguably, each of (i) and (ii) is a reasonable approach to take, but they may yield different

[1] This assumption is natural in a regulatory setting such as the FRTB, where risk measures are heavily used; see also the above mentioned references. Other ways to aggregate risk models, such as averaging, max-min, smooth aggregation (Klibanoff et al. (2005)) and anti-conservative (e.g., best-case) approaches, may be suitable in different contexts, and they are outside the scope of the current paper.

risk evaluations. We shall call (i) the *worst-case risk (WR) approach*, and (ii) the *model aggregation (MA) approach*. There are two obvious advantages of the MA approach: We obtain a robust model which is useful for analysis and simulation, thus answering the motivating question above, and the procedure applies for generic risk measures, not only a specific one. Other less obvious, but important, advantages of the MA approach will be revealed through this paper. The model F^* is robust in two senses: First, it is more conservative than any models in $\mathcal{F}$; second, it applies to a wide range of risk measures or decision criteria.

At this point, we have not yet specified how the robust distributional model F^* may be obtained in the MA approach (ii). For this purpose, we need an order relation, often consistent with the risk measure ρ used by the risk analyst. We will describe some natural choices of partial orders, in particular, first- and second-order stochastic dominance, in Section 2.1.

Our main objective is a comprehensive theory on the two approaches of robust risk evaluation, with a focus on the newly introduced MA approach. The following questions naturally arise.

Q1. What are the advantages and disadvantages of the MA approach in contrast to the WR approach, in addition to the points mentioned above?

Q2. What are theoretical and computational properties of the MA approach in distributionally robust optimization?

Q3. Which risk measures yield equivalent robust risk evaluation results via the MA and WR approaches and how are they used in regulatory practice?

Q4. How is the MA approach implemented in common settings of uncertainty, optimization, and real-data applications?

We will answer the four questions above by means of several novel theoretical results. Our main contributions can be explained as follows. After introducing partial orders and risk measures in Section 2, we present a rigorous formulation of the MA and WR approaches for risk evaluation and distributionally robust optimization in Section 3. These optimization problems will be called WR or MA optimization. Our new method is related to stochastic optimization of risk measures (e.g., Dentcheva and Ruszczynski (2004), Ruszczyński and Shapiro (2006), Shapiro et al. (2021)); see our discussion in Section 3.

Features of MA risk evaluation will be discussed in Section 4 and MA robust optimization will be studied in Section 5. We show convenient properties of the MA approach in risk evaluation and optimization. In particular, the MA risk evaluation remains convex when the risk measure is convex, and the MA robust optimization admits a convex program reformulation under suitable conditions. This answers Q2, and also Q1 partially.

We establish in Section 7 that the property of equivalence in model aggregation characterizes VaR and ES among very general classes of risk measures. The equivalence property identifies for which risk measures the two approaches can be converted to each other. Through these results, which require long technical proofs, the rich literature of robust risk evaluation and optimization, popular in operations research,[2] is connected to that of the axiomatic theory of risk preferences, popular in decision theory,[3] for the first time. Our results contribute to the latter literature by offering new axiomatizations of both VaR and ES which are important issues in risk management in themselves.[4] These results answer question Q3 above.

We address two settings of uncertainty, those generated by Wasserstein metrics and those generated by moment information in Section 6. We illustrate that the MA approach leads to closed-form robust distributional models in these settings, being easy to apply and computationally feasible. Based on a new result on dimension reduction for Wasserstein balls (Theorem 5), we show that the MA approach can conveniently handle multivariate Wasserstein uncertainty in the setting of portfolio selection. Section 8 contains two applications of worst-case risk evaluation and portfolio selection under uncertainty using real financial data. These two sections answer Q4.

Finally, advantages and limitations of the MA approach, as well as directions for future work, are summarized and discussed in Section 9, which also contains a preliminary discussion on aggregating multivariate risk models, in contrast to the univariate risk models treated throughout the paper. These discussions address Q1 at a high level.

In the main text of the paper, we focus on the set of distributions with finite mean to make our analysis concise and managerial insights clear. More general choices of the space of distributions are treated in the appendices, which also contain technical proofs of all results.

2 Preliminaries and standing notation

We first introduce some notation. Let $(\Omega, \mathcal{B}, \mathbb{P})$ be a nonatomic probability space. For $d \in \mathbb{N}$, let $\mathfrak{B}(\mathbb{R}^d)$ be the Borel σ-field on $\mathbb{R}^d$. A random vector $\mathbf{X}$ is a measurable mapping from $(\Omega, \mathcal{B})$

[2] In addition to the literature on portfolio optimization, robust risk evaluation and optimization also broadly exist in other applications of operations research; see Wiesemann et al. (2014), Esfahani and Kuhn (2018), Blanchet et al. (2019) and Embrechts et al. (2022) for a small specimen.

[3] For developments on axiomatic studies in decision theory, see e.g., Klibanoff et al. (2005), Maccheroni et al. (2006) and Cerreia-Vioglio et al. (2021). Axiomatic theory of risk measures have also been an active topic in quantitative finance since the seminal work of Artzner et al. (1999); see Föllmer and Schied (2016) for a comprehensive treatment.

[4] In particular, Chambers (2009) obtained an axiomatization of VaR and Wang and Zitikis (2021) obtained an axiomatization of ES; see also Remarks 5 and 6 for other axiomatizations of VaR and ES.

to $(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d))$; a random variable is a random vector with $d = 1$. Denote by $\mu_{\mathbf{X}} := \mathbb{P} \circ \mathbf{X}^{-1}$ the probability distribution induced by a random vector $\mathbf{X}$ under $\mathbb{P}$, where $\mathbf{X}^{-1}$ is the inverse image of $\mathbf{X}$. Denote by $F_{\mathbf{X}}$ the cumulative distribution function (cdf) of $\mathbf{X}$ under $\mathbb{P}$, i.e., $F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}(\mathbf{X} \leqslant \mathbf{x})$ for $\mathbf{x} \in \mathbb{R}^d$, where the inequality is component-wise. We will take cdfs $F_{\mathbf{X}}$, identified with distributions $\mu_{\mathbf{X}}$, as the main research object of the entire paper. We use $\delta_{\mathbf{t}}$ to represent the point-mass at $\mathbf{t} \in \mathbb{R}^d$. Let L^1 be the space of all integrable random variables on $(\Omega, \mathcal{B}, \mathbb{P})$, where almost surely equal random variables are treated as identical. Denote by $\mathcal{M}_1$ the set of cdfs of all random variables in L^1, i.e., $\mathcal{M}_1$ is the set of all cdfs F satisfying $\int_{\mathbb{R}} |x| \mathrm{d}F(x) < \infty$. The mean of a random variable X under $\mathbb{P}$ is written as $\mathbb{E}[X] = \mathfrak{m}(F_X)$, where $\mathbb{E}$ is defined on L^1 and $\mathfrak{m}$ is defined on $\mathcal{M}_1$. We denote by Δ_d the standard simplex $\{\boldsymbol{\lambda} \in [0,1]^d : \sum_{i=1}^d \lambda_i = 1\}$ in $\mathbb{R}^d$ and by $[d] = \{1, \dots, d\}$.

2.1 Stochastic orders and lattices

For any set of cdfs $\mathcal{M}$, let $\preceq$ be a partial order on $\mathcal{M}$, and $(\mathcal{M}, \preceq)$ is called a partially ordered set. The relevant tool is the lattice theory which we collect in Appendix C, and here we only present a basic result needed to understand our main ideas. The most commonly used partial orders in finance and economics are the first-order stochastic dominance $\preceq_1$ and the increasing convex order $\preceq_2$, defined as, for $F, G \in \mathcal{M}$,[5]

(a) $F \preceq_1 G$ if $\int u \mathrm{d}F \leqslant \int u \mathrm{d}G$ for all increasing functions u;

(b) $F \preceq_2 G$ if $\int u \mathrm{d}F \leqslant \int u \mathrm{d}G$ for all increasing convex functions u.

Other useful equivalent definitions of $\preceq_1$ and $\preceq_2$ are put in Appendix C. To build a robust distributional model, we need to define the supremum of a set $\mathcal{F} \subseteq \mathcal{M}$. Throughout, $\mathcal{F}$ is assumed nonempty. For a partial ordered set $(\mathcal{M}, \preceq)$ and $\mathcal{F} \subseteq \mathcal{M}$, let

$$\underline{U}(\mathcal{F}) = \{H \in U(\mathcal{F}) : H \preceq G,\ \forall G \in U(\mathcal{F})\} \ \text{ with } \ U(\mathcal{F}) = \{G \in \mathcal{M} : F \preceq G,\ \forall F \in \mathcal{F}\}.$$

Note that $\underline{U}(\mathcal{F})$ is by definition either a singleton or an empty set. If $\underline{U}(\mathcal{F})$ is not empty, then the supremum of $\mathcal{F}$, denoted by $\bigvee \mathcal{F}$, is defined as the unique element of $\underline{U}(\mathcal{F})$; otherwise it is undefined.[6] If $U(\mathcal{F})$ is not empty, we say that $\mathcal{F}$ is *bounded from above with respect to* $\preceq$ ($\preceq$-

[5] Note that we treat F and G as loss cdfs instead of wealth cdfs, and hence a larger element in $\preceq_1$ or $\preceq_2$ means higher risk, see e.g., Shaked and Shanthikumar (2007). For a cdf F, denote by $\widetilde{F}$ the cdf of $-X$ where X follows F. Up to a sign flip, increasing convex order (for loss distributions) is economically equivalent to second-order stochastic dominance (for gain distributions) in the sense that $F \succeq_2 G$ if and only if $\widetilde{F} \preceq_{\mathrm{ssd}} \widetilde{G}$, where $\preceq_{\mathrm{ssd}}$ is defined via (b) by changing increasing convex functions to increasing concave functions; see Lemma EC.1 in Appendix B .

[6] Note that this definition is analogous to the definition of supremum for a subset A of the real line: $\sup A$ is defined as the smallest real number that dominates every element in A, if A is bounded from above; otherwise the supremum of A does not exist in $\mathbb{R}$ (and by default set to ∞).

bounded, for short). The supremum does not always exist, but for the two choices of ordered sets $(\mathcal{M}_1, \preceq_1)$ and $(\mathcal{M}_1, \preceq_2)$ that we consider in the main paper, this does not create any problem; see e.g., Kertz and Rösler (2000) and Müller and Scarsini (2006) for the lattice structure of cdfs with $\preceq_1$ and $\preceq_2$. The two cases of $\bigvee \mathcal{F}$ for $\preceq_1$ and $\preceq_2$ admit explicit formulas, given in Proposition 1 in Section 4.

2.2 Risk measures

In the classic framework of Artzner et al. (1999) and Föllmer and Schied (2016), a risk measure is traditionally defined as a mapping from a set $\mathcal{X}$ of random losses to $\mathbb{R}$. Denote by $\mathcal{M}$ the set of cdfs of random variables in $\mathcal{X}$. For a partial order $\preceq$ on $\mathcal{M}$, a natural interpretation of $F \preceq G$ is that G is riskier than F according to $\preceq$. A risk measure $\rho : \mathcal{M} \to \mathbb{R}$ is $\preceq$-*consistent* if $\rho(F) \leqslant \rho(G)$ for all $F, G \in \mathcal{M}$ with $F \preceq G$.

Definition 1. A *distribution based risk measure* is a mapping $\rho : \mathcal{M} \to \mathbb{R}$ satisfying $\preceq_1$-consistency. For such ρ, its associated *random-variable based risk measure* is $\widetilde{\rho} : \mathcal{X} \to \mathbb{R}$ given by $\widetilde{\rho}(X) = \rho(F_X)$. Both ρ and $\widetilde{\rho}$ will be called risk measures in this paper.

Note that ρ is $\preceq_1$-consistent if and only if $\widetilde{\rho}$ is monotone (i.e., $\widetilde{\rho}(X) \leqslant \widetilde{\rho}(Y)$ when $X \leqslant Y$). The random-variable based risk measure in Definition 1 satisfies *law-invariance* (i.e., $\widetilde{\rho}(X) = \widetilde{\rho}(Y)$ whenever $F_X = F_Y$). There exists a one-to-one correspondence between $\rho : \mathcal{M} \to \mathbb{R}$ and $\tilde{\rho} : \mathcal{X} \to \mathbb{R}$ satisfying law-invariance; see e.g., Proposition 1 of Delage et al. (2019). We will choose $\mathcal{X} = L^1$ in the main part of the paper, so that the two partial orders $\preceq_1$ and $\preceq_2$ both behave well.[7] For a better exposition of distributional uncertainty, we will present ideas and results mainly using ρ instead of $\widetilde{\rho}$.

The two most popular and important risk measures in financial practice, VaR and ES, are both law-invariant. The risk measure VaR at level $\alpha \in (0,1)$ is the functional $\mathrm{VaR}_\alpha : \mathcal{M}_1 \to \mathbb{R}$ defined by

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

which is the left α-quantile of a cdf. The risk measure ES at level $\alpha \in [0,1)$ is the functional $\mathrm{ES}_\alpha : \mathcal{M}_1 \to \mathbb{R}$ defined by

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

[7] In particular, it is well known that $\preceq_2$ is closely related to mean-preserving spreads of Rothschild and Stiglitz (1970), and a finite mean is essential for such a connection. On the other hand, $\preceq_1$ fits well in any space of random variables or cdfs.

and in particular, $\mathrm{ES}_0(F) = \mathfrak{m}(F)$. We can also define VaR_0, VaR_1 and ES_1, which are not finite-valued on $\mathcal{M}_1$; see Appendix D. Note that in addition to the $\preceq_1$-consistency of VaR and ES, ES also satisfies the $\preceq_2$-consistency.

3 Introducing the MA approach

We describe the two approaches for robust risk evaluation, the primary objects of this paper. For a risk measure $\rho : \mathcal{M} \to \mathbb{R}$ and an uncertainty set $\mathcal{F} \subseteq \mathcal{M}$, a common way to obtain a robust risk evaluation is to calculate the following worst-case risk measure

$$\mathrm{WR}: \qquad \rho^{\mathrm{WR}}(\mathcal{F}) = \sup_{F \in \mathcal{F}} \rho(F). \tag{1}$$

The value in (1) is called the *WR robust ρ value*, and it has been widely studied in the literature; some references are mentioned in the introduction. Next, we propose a new method of robust risk evaluation, that is, assuming that the supremum $\bigvee \mathcal{F}$ with respect to $\preceq$ exists,

$$\mathrm{MA}: \qquad \rho^{\mathrm{MA}}(\mathcal{F}) = \rho\left(\bigvee \mathcal{F}\right), \tag{2}$$

and $\rho^{\mathrm{MA}}(\mathcal{F}) = \infty$ if $\mathcal{F}$ is not bounded from above. The value in (2) is called the *$\preceq$-MA robust ρ value* ("$\preceq$-" will be omitted if the order is clear from the context). In the main text of the paper, $\bigvee \mathcal{F}$ exists for all $\mathcal{F}$ bounded from above, and hence, ρ^{MA} is always well-defined. In case that $\bigvee \mathcal{F}$ may not exist, (2) needs to be modified as in Appendix C.

The idea of the MA approach can be described in two steps: First, take the supremum $\bigvee \mathcal{F}$ of the uncertainty set $\mathcal{F}$ as the robust distribution, and second, calculate the value of the risk measure of the robust distribution. The robust distribution $\bigvee \mathcal{F}$ obtained in the first step can be used for any risk measure. If, in addition, the risk measure ρ is $\preceq$-consistent, then the MA approach produces a larger robust risk value than the WR approach, that is, for any $\mathcal{F} \subseteq \mathcal{M}$,

$$\rho^{\mathrm{WR}}(\mathcal{F}) \leqslant \rho^{\mathrm{MA}}(\mathcal{F}), \tag{3}$$

since $\preceq$-consistency implies $\rho(\bigvee \mathcal{F}) \geqslant \rho(F)$ for all $F \in \mathcal{F}$. The MA approach can be implemented even in case no risk measure is involved (thus skipping the second step above), as the model $\bigvee \mathcal{F}$ is ready to use without a specification of any specific objective.

In the sequel, we will focus on $\preceq_1$ and $\preceq_2$. For a simpler notation, we write MA_1 when $\preceq$ is specified as $\preceq_1$, and MA_2 is similar. For these two stochastic orders, the explicit forms of $\bigvee \mathcal{F}$

are obtained in Section 4. It is also worth noting that if ρ is consistent with more than one partial orders, then the MA approach with a stronger partial order leads to a higher risk evaluation. For instance, if ρ is both $\preceq_1$-consistent and $\preceq_2$-consistent, then $\rho^{\mathrm{WR}}(\mathcal{F}) \leqslant \rho^{\mathrm{MA}_2}(\mathcal{F}) \leqslant \rho^{\mathrm{MA}_1}(\mathcal{F})$ because any $(\mathcal{M}, \preceq_1)$-upper bound on $\mathcal{F}$ is also an $(\mathcal{M}, \preceq_2)$-upper bound on $\mathcal{F}$.

Using the WR and MA approaches, two types of distributionally robust optimization problems arise:

$$\min_{\mathbf{a}\in A} \ \rho^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f}) \quad \text{and} \quad \min_{\mathbf{a}\in A} \ \rho^{\mathrm{MA}}(\mathcal{F}_{\mathbf{a},f}), \tag{4}$$

where A is a set of possible actions, $f : A \times \mathbb{R}^d \to \mathbb{R}$ is a loss function, $\mathcal{F}$ is a set of cdfs on $\mathbb{R}^d$ and

$$\mathcal{F}_{\mathbf{a},f} = \left\{ F_{f(\mathbf{a},\mathbf{X})} : F_{\mathbf{X}} \in \mathcal{F} \right\}. \tag{5}$$

The set $\mathcal{F}_{\mathbf{a},f}$ consists all univariate cdfs of $f(\mathbf{a}, \mathbf{X})$ where $\mathbf{X}$ has cdf in $\mathcal{F}$. For instance, by choosing $A \subseteq \mathbb{R}^d$ and $f(\mathbf{a}, \mathbf{x}) = \mathbf{a}^\top \mathbf{x}$, one arrives at the setting of robust portfolio selection, where $\mathbf{a}$ represents the vector of portfolio weights and $\mathbf{x}$ represents the vector of losses from individual assets. The WR robust optimization problem in (4) is equivalent to a minimax problem:

$$\min_{\mathbf{a}\in A} \ \sup_{F\in\mathcal{F}} \widetilde{\rho}^{F}(f(\mathbf{a}, \mathbf{X})), \tag{6}$$

where $\widetilde{\rho}^{F}(f(\mathbf{a}, \mathbf{X}))$ represents the value of $\widetilde{\rho}(f(\mathbf{a}, \mathbf{X}))$ when $\mathbf{X}$ has the cdf F. If ρ is $\preceq$-consistent, then the MA robust optimization problem in (4) can be converted to a stochastic program with partial order $\preceq$ constraints:

$$\min_{\mathbf{a}\in A} \ \inf_{H} \rho(H) \qquad \text{s.t. } G \preceq H, \ \ \forall G \in \mathcal{F}_{\mathbf{a},f}. \tag{7}$$

The above problem with $\preceq$ being $\preceq_2$ will be studied in Section 5. Section 6 is dedicated to the portfolio selection problem, where two specific settings of uncertainty will be considered.

Comparison: Optimization with stochastic dominance. Dentcheva and Ruszczynski (2004) introduced an optimization problem with stochastic dominance constraint. Adapting to our notation, and focusing on $\preceq_2$, their model can be described as (see Appendix B for the details on this equivalent representation)

$$\min_{\mathbf{a}\in A} \ \ \mathbb{E}[f(\mathbf{a}, \mathbf{X})] \qquad \text{s.t. } F_{g_i(\mathbf{a},\mathbf{X})} \preceq_2 F_i, \ \ i \in [m], \tag{8}$$

where $\mathbf{X}$ has a fixed cdf, $f, g_1, \ldots, g_m$ are fixed functions, and $F_1, \ldots, F_m \in \mathcal{M}_1$. Note that if we replace $\mathbb{E}$ by $\widetilde{\rho}$ and set $g_1 = \cdots = g_m = f$, then the problem is

$$\min_{\mathbf{a}\in A} \ \widetilde{\rho}(f(\mathbf{a},\mathbf{X})) \qquad \text{s.t. } F_{f(\mathbf{a},\mathbf{X})} \preceq_2 F_i, \ \ i \in [m]. \tag{9}$$

The problem (9) is similar to our MA problem (7), but there are a few essential differences. First, for fixed $\mathbf{a} \in A$, the cdf of $f(\mathbf{a},\mathbf{X})$ is fixed in (9), whereas (7) searches for a robust model for $f(\mathbf{a},\mathbf{X})$ over the uncertainty set $\mathcal{F}_{\mathbf{a},f}$. Second, the direction of stochastic dominance is flipped, as our H dominates every G in $\mathcal{F}_{\mathbf{a},f}$ and their $F_{f(\mathbf{a},\mathbf{X})}$ is dominated by every F_i. Note that the interpretation of stochastic dominance is very different here: (7) looks at risks larger in $\preceq_2$ (riskier) because our objective is robust optimization, whereas (9) looks at risks smaller in $\preceq_2$ (safer) because of risk constraint. In Dentcheva and Ruszczynski (2003), the following problem has been considered: $\min_{X\in\mathcal{C}} \widetilde{\rho}(X)$ s.t. $F_X \preceq_2 F_i, \ i \in [m]$, where $\mathcal{C}$ is a set of random variables. When ρ is $\preceq_2$-consistent, our MA risk evaluation problem can be written as $\min_{X\in L^1} \widetilde{\rho}(X)$ s.t. $F_i \preceq_2 F_X, \ i \in [m]$, which is similar to the model of Dentcheva and Ruszczynski (2003); see the recent work of Dentcheva (2023) for a related two-stage optimization with stochastic dominance constraints.

Remark 1. In this paper, we focus on risk measures taking real values. In some applications, risk measures may be multi-valued or set-valued (e.g., Embrechts and Puccetti (2006), Hamel and Heyde (2010), Hamel et al. (2011)). For such risk measures, the MA approach, together with multivariate stochastic order, can also be applied.

4 MA approach in risk evaluation

In this section, we study properties of the MA risk evaluation by focusing on $\preceq_1$ and $\preceq_2$.

4.1 Computing the robust model

We use $\bigvee_1 \mathcal{F}$ and $\bigvee_2 \mathcal{F}$ to represent the supremum of the uncertainty set $\mathcal{F}$ on the ordered set $(\mathcal{M}_1, \preceq_1)$ and $(\mathcal{M}_1, \preceq_2)$, respectively, and π_{F_X} represents the *integrated survival function* of $X \in L^1$, defined as

$$\pi_{F_X}(x) = \int_x^\infty (1 - F_X(t))\mathrm{d}t = \mathbb{E}[(X-x)_+], \quad x \in \mathbb{R}, \tag{10}$$

where $x_+ = \max\{0, x\}$ for $x \in \mathbb{R}$. It is straightforward from (10) that a simple relationship between the integrated survival function π_F and the cdf F is $F = 1 + (\pi_F)'_+$, where $(\pi_F)'_+$ is the right derivative of π_F. The left quantile function of $F \in \mathcal{M}_1$ is defined by $F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geqslant$

$\alpha\}$ for $\alpha \in (0,1]$, which is $\mathrm{VaR}_\alpha(F)$ when $\alpha \in (0,1)$. The functions π_F and F^{-1} will be used throughout the paper. The supremum and infimum over functions are understood point-wise.

Proposition 1. *(a) For a set $\mathcal{F} \subseteq \mathcal{M}_1$ that is $\preceq_1$-bounded, we have $\bigvee_1 \mathcal{F} = \inf_{F\in\mathcal{F}} F$*[8] *and the left quantile function of $\bigvee_1 \mathcal{F}$ is $\sup_{F\in\mathcal{F}} F^{-1}$.*

(b) For a set $\mathcal{F} \subseteq \mathcal{M}_1$ that is $\preceq_2$-bounded, we have

$$\pi_{\bigvee_2 \mathcal{F}} = \sup_{F\in\mathcal{F}} \pi_F, \text{ and thus } \bigvee_2 \mathcal{F} = 1 + \left(\sup_{F\in\mathcal{F}} \pi_F\right)'_+, \tag{11}$$

where g'_+ denotes the right derivative of g.

Remark 2. Under the condition of Proposition 1 (b), the function $\bigvee_2 \mathcal{F} = 1 + (\sup_{F\in\mathcal{F}} \pi_F)'_+$ is a well-defined cdf by noting that π_F is a decreasing convex function, and thus, its right derivative is well-defined, nonnegative, and right continuous. More details can be found in the proof of Proposition EC.1.

By Proposition 1, for a risk measure ρ, the evaluation of MA_1 is equivalent to applying ρ to the distribution with a worst-case quantile function. Similarly, the evaluation of MA_2 is equivalent to applying ρ to the distribution with a worst-case upper partial moment given by (11). Worst-case quanitle and worst-case upper partial moment functions have wide range of applications in optimization; see e.g., El Ghaoui et al. (2003), Lo (1987), Natarajan et al. (2010) and Cowell (2011). To compute $\pi_{\bigvee_2 \mathcal{F}}$ numerically, as a decreasing convex function, $\pi_{\bigvee_2 \mathcal{F}}$ can be well approximated by a piece-wise linear function with finitely many pieces (which requires computing it at finitely many points). The next result concerns convexity of the uncertainty set.

Proposition 2. *Suppose that $i \in \{1,2\}$. For $\mathcal{F} \subseteq \mathcal{M}_1$, we have $\bigvee_i \mathrm{conv}\mathcal{F} = \bigvee_i \mathcal{F}$, where $\mathrm{conv}\mathcal{F}$ is the convex hull of $\mathcal{F}$.*

Proposition 2 illustrates that for the MA_1 and MA_2 approaches, one can convert freely between any uncertainty set and its convex hull. For WR risk evaluation, this does not hold in general.

4.2 VaR and ES

Next we discuss the MA and WR approaches applied to VaR and ES, and this will help us understand the inequality (3). The case of VaR, coupled with the partial order $\preceq_1$, is simple. By

[8]Note that the infimum of upper semicontinuous functions $F \in \mathcal{F}$ is again upper semicontinuous and thus a valid cdf when $\mathcal{F}$ is $\preceq_1$-bounded.

Proposition 1, for $\alpha \in (0,1)$ and any $\mathcal{F}$ that is $\preceq_1$-bounded, $\mathrm{VaR}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \mathrm{VaR}_\alpha^{\mathrm{MA}_1}(\mathcal{F})$, and thus (3) holds as an equality in this specific setting; this result will be collected in Theorem 1 below.

The case of ES is more illuminating. Note that ES is consistent with respect to both $\preceq_1$ and $\preceq_2$. First, we consider the MA approach with $\preceq_1$. Since

$$\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \frac{1}{1-\alpha} \sup_{F\in\mathcal{F}} \int_\alpha^1 F^{-1}(s)\mathrm{d}s \leqslant \frac{1}{1-\alpha} \int_\alpha^1 \sup_{F\in\mathcal{F}} F^{-1}(s)\mathrm{d}s = \mathrm{ES}_\alpha^{\mathrm{MA}_1}(\mathcal{F}),$$

for (3) to hold as an equality, one needs to exchange the order of a supremum and an integral. Such an exchange, if legitimate, means that there exists $F \in \mathcal{F}$ such that $F^{-1}(s) \geqslant G^{-1}(s)$ for all $G \in \mathcal{F}$ and $s \in (\alpha, 1)$, which is a quite strong assumption unlikely to hold in applications.

Next, we consider the MA approach for ES with $\preceq_2$. Recall a representation of ES_α for $\alpha \in (0,1)$ in the celebrated work of Rockafellar and Uryasev (2002) and Pflug (2000), that is,

$$\mathrm{ES}_\alpha(F) = \min_{x\in\mathbb{R}} \left\{ x + \frac{1}{1-\alpha}\pi_F(x) \right\}, \quad F \in \mathcal{M}_1. \tag{12}$$

Using (12), we obtain the WR robust ES value, that is,

$$\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \sup_{F\in\mathcal{F}} \mathrm{ES}_\alpha(F) = \sup_{F\in\mathcal{F}} \min_{x\in\mathbb{R}} \left\{ x + \frac{1}{1-\alpha}\pi_F(x) \right\}. \tag{13}$$

On the other hand, the $\preceq_2$-MA robust ES value can also be calculated using (12) and (11) in Proposition 1, that is,

$$\mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}) = \mathrm{ES}_\alpha\left(\bigvee_2 \mathcal{F}\right) = \min_{x\in\mathbb{R}} \sup_{F\in\mathcal{F}} \left\{ x + \frac{1}{1-\alpha}\pi_F(x) \right\}, \tag{14}$$

where the second equality follows from (12) and $\pi_{\bigvee_2 \mathcal{F}}(x) = \sup_{F\in\mathcal{F}} \pi_F(x)$ by Proposition 1 (ii). The formulas (13) and (14) imply that the WR and MA robust ES values can be seen as, respectively, the maximin and the minimax of the same bivariate objective function. This observation immediately leads to

$$\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) \leqslant \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}), \text{ and equality holds if a minimax theorem holds.} \tag{15}$$

Therefore, although (3) is generally not an equality, it may be an equality for ES_α and $\preceq_2$ under certain conditions on $\mathcal{F}$. In particular, as shown by Zhu and Fukushima (2009), if $\mathcal{F}$ is a convex polytope (see Section 7 for a definition) or a compact convex set of discrete cdfs, then (15) becomes an equality. In the following theorem, we establish a more general sufficient condition to make (15) an equality, where $\mathrm{ES}_0 = \mathfrak{m}$ and ES_α for $\alpha \in (0,1)$ are treated separately. We also collect the corresponding result for VaR_α discussed above. Recall that ρ^{WR} and ρ^{MA} are defined in (1) and (2) for any risk measure ρ.

Theorem 1. *Suppose that $\mathcal{F} \subseteq \mathcal{M}_1$.*

(a) If $\sup_{F\in\mathcal{F}} \int_{\mathbb{R}} (x-y)_+ \mathrm{d}F(y) \to 0$ as $x \to -\infty$, then $\mathfrak{m}^{\mathrm{WR}}(\mathcal{F}) = \mathfrak{m}^{\mathrm{MA}_2}(\mathcal{F})$.

(b) For $\alpha \in (0,1)$, if $\mathcal{F}$ is convex and $\preceq_2$-bounded, then $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F})$.

(c) For $\alpha \in (0,1)$, if $\mathcal{F}$ is $\preceq_1$-bounded, then $\mathrm{VaR}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \mathrm{VaR}_\alpha^{\mathrm{MA}_1}(\mathcal{F})$.

The most useful part of Theorem 1 is (b), which offers a simple condition under which the WR robust ES value can be obtained by implementing the MA approach. This result generalizes Theorems 1 and 2 of Zhu and Fukushima (2009) where the set $\mathcal{F}$ is a convex polytope and a compact convex set of discrete cdfs, respectively. Without convexity of $\mathcal{F}$, for $\alpha \in (0,1)$, $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F})$ may not hold, as illustrated by the following example.

Example 1. Let $\alpha \in (0,1)$. Let $\varepsilon = (1-\alpha)/2$, $F_1 = \delta_0$ and $F_2 = (1-\varepsilon)\delta_{-1/(1-\varepsilon)-1} + \varepsilon\delta_{1/\varepsilon}$, where δ_t represents the point-mass at $t \in \mathbb{R}$. By computing $\max\{\pi_{F_1}, \pi_{F_2}\}$, we get $\bigvee_2\{F_1, F_2\} = (1-\varepsilon)\delta_{-1/(1-\varepsilon)} + \varepsilon\delta_{1/\varepsilon}$ and

$$\mathrm{ES}_\alpha\left(\bigvee_2 \{F_1, F_2\}\right) = \frac{1}{2}\left(\frac{1}{\varepsilon} - \frac{1}{1-\varepsilon}\right) > \frac{1}{2}\left(\frac{1}{\varepsilon} - \frac{2-\varepsilon}{1-\varepsilon}\right)_+ = \max\left\{\mathrm{ES}_\alpha(F_1), \mathrm{ES}_\alpha(F_2)\right\}.$$

Hence, $\mathrm{ES}_\alpha^{\mathrm{WR}}(\{F_1, F_2\}) < \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\{F_1, F_2\})$.

The conditions on $\mathcal{F}$ in (a) and (b) of Theorem 1 do not imply each other. The following example shows that $\mathfrak{m}^{\mathrm{WR}}(\mathcal{F}) = \mathfrak{m}^{\mathrm{MA}_2}(\mathcal{F})$ may not hold in case $\mathcal{F}$ does not satisfy the condition in (a) and satisfy the condition in (b).

Example 2. For $n \in \mathbb{N}$, let $F_n = (1/n)\delta_{-n} + (1-1/n)\delta_0$, and denote by $\mathcal{F}$ the convex hull of $\{F_n\}_{n\in\mathbb{N}}$. By computing $\max\{\pi_{F_n}, n \in \mathbb{N}\}$, we have $\bigvee_2 \mathcal{F} = \bigvee_2\{F_n\}_{n\in\mathbb{N}} = \delta_0$. Note that $\mathfrak{m}(F) = -1$ for any $F \in \mathcal{F}$. Hence, $\mathfrak{m}(\bigvee_2 \mathcal{F}) = 0 > -1 = \sup_{F\in\mathcal{F}} \mathfrak{m}(F)$, that is, $\mathfrak{m}^{\mathrm{WR}}(\mathcal{F}) < \mathfrak{m}^{\mathrm{MA}_2}(\mathcal{F})$.

4.3 Convexity and other properties

We first introduce some standard properties of a risk measure $\rho : \mathcal{M}_1 \to \mathbb{R}$ and its associated $\widetilde{\rho} : L^1 \to \mathbb{R}$. *Translation invariance*: $\widetilde{\rho}(X + c) = \widetilde{\rho}(X) + c$ for any $c \in \mathbb{R}$ and $X \in L^1$. *Positive homogeneity*: $\widetilde{\rho}(\lambda X) = \lambda\widetilde{\rho}(X)$ for any $\lambda > 0$ and $X \in L^1$. *Convexity*: $\widetilde{\rho}(\lambda X + (1-\lambda)Y) \leqslant \lambda\widetilde{\rho}(X) +$

$(1-\lambda)\widetilde{\rho}(Y)$ for any $\lambda \in [0,1]$ and $X, Y \in L^1$. *Lower semicontinuity*: $\liminf_{n\to\infty} \widetilde{\rho}(X_n) \geqslant \widetilde{\rho}(X)$ if $X_n, X \in L^1$ for all n and $X_n \xrightarrow{\text{d}} X$ as $n \to \infty$, where $\xrightarrow{\text{d}}$ denotes convergence in distribution.[9] *Comonotonic additivity*: $\widetilde{\rho}(X+Y) = \widetilde{\rho}(X) + \widetilde{\rho}(Y)$ for any $X, Y \in L^1$ that are comonotonic.[10] A risk measure is *coherent*, as defined by Artzner et al. (1999), if it satisfies translation invariance, positive homogeneity, and convexity (also monotonicity, which is assumed in Definition 1). All the properties are defined for both ρ and $\widetilde{\rho}$.

It is well known that VaR_α and ES_α, $\alpha \in (0,1)$ satisfy translation invariance, positive homogeneity, lower semicontinuity, comonotonic additivity, and ES_α further satisfies convexity. Translation invariance, positive homogeneity and convexity are standard properties with interpretations extensively discussed by Artzner et al. (1999) and Föllmer and Schied (2016). Lower semicontinuity, called the prudence axiom by Wang and Zitikis (2021), means that if the loss cdf is modeled using a truthful approximation, then the approximated risk model should not underreport the capital requirement. Comonotonic additivity is popular in both the literature of decision theory (e.g., Yaari (1987) and Schmeidler (1989)) and that of risk measures (e.g., Kusuoka (2001)). A coherent risk measure on $\mathcal{M}_1$, including ES, is automatically consistent with both $\preceq_1$ and $\preceq_2$; see e.g., Leitner (2005), Föllmer and Schied (2016) and Shapiro et al. (2021).

Next we formulate uncertainty for $\widetilde{\rho}$ with different probabilities based on $\rho : \mathcal{M}_1 \to \mathbb{R}$. Let $\mathcal{P}$ be the set of all probability measures on $(\Omega, \mathcal{B})$ absolutely continuous with respect to $\mathbb{P}$. For $Q \in \mathcal{P}$, define $L^1(\Omega, \mathcal{B}, Q)$ as the set of random variables (in the sense of $\mathbb{P}$-almost sure equivalence class) that have a finite first moment under Q. Note that since $Q \in \mathcal{P}$, if two random variables are equal $\mathbb{P}$-almost surely, then they are equal Q-almost surely. Define $\widetilde{\rho}^Q(X) = \rho(F_X^Q)$, $X \in L^1(\Omega, \mathcal{B}, Q)$, where F_X^Q denotes the cdf of X under $Q \in \mathcal{P}$. In particular, we have $\widetilde{\rho}^{\mathbb{P}}(X) = \widetilde{\rho}(X) = \rho(F_X)$ for all $X \in L^1$. For $\mathcal{Q} \subseteq \mathcal{P}$, denote by $L = \bigcap_{Q \in \mathcal{Q}} L^1(\Omega, \mathcal{B}, Q)$ and we use $\mathcal{F}_{X|\mathcal{Q}}$ to represent the set of all possible cdfs of $X \in L$ under the uncertainty set, i.e., $\mathcal{F}_{X|\mathcal{Q}} = \{F_X^Q : Q \in \mathcal{Q}\}$.

In this setting, we study the properties of risk measures on L via WR approach

$$\text{WR}: \qquad \widetilde{\rho}^{\mathrm{WR}}(X) = \rho^{\mathrm{WR}}\left(\mathcal{F}_{X|\mathcal{Q}}\right) = \sup_{F \in \mathcal{F}_{X|\mathcal{Q}}} \rho(F) = \sup_{Q \in \mathcal{Q}} \widetilde{\rho}^Q(X),$$

[9] Convergence in distribution corresponds to weak convergence on $\mathcal{M}_1$. Note that this lower semicontinuity is different from L^1-lower semicontinuity commonly used in the literature of risk measures (e.g., Föllmer and Schied (2016)).

[10] Random variables X and Y are *comonotonic* if there exists $\Omega_0 \in \mathcal{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that for all $\omega, \omega' \in \Omega_0$,

$$(X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) \geqslant 0.$$

and via MA_i approach for $i = 1, 2$,

$$\text{MA}: \quad \widetilde{\rho}^{\mathrm{MA}_i}(X) = \rho^{\mathrm{MA}_i}\left(\mathcal{F}_{X|\mathcal{Q}}\right) = \begin{cases} \rho\left(\bigvee_i \mathcal{F}_{X|\mathcal{Q}}\right), & \text{if } \mathcal{F}_{X|\mathcal{Q}} \text{ is } \preceq_i\text{-bounded;} \\ \infty, & \text{otherwise.} \end{cases}$$

The two mappings $\widetilde{\rho}^{\mathrm{WR}}$ and $\widetilde{\rho}^{\mathrm{MA}_i}$ are well defined on the set $L = \bigcap_{Q\in\mathcal{Q}} L^1(\Omega, \mathcal{B}, Q)$. This set always contains, for instance, all bounded random variables in L^1. The next result gives properties of $\widetilde{\rho}^{\mathrm{WR}}$ and $\widetilde{\rho}^{\mathrm{MA}_i}$ based on those of $\widetilde{\rho}$.

Theorem 2. *Let $\widetilde{\rho} : L^1 \to \mathbb{R}$ be a risk measure, $\mathcal{Q} \subseteq \mathcal{P}$, and $L = \bigcap_{Q\in\mathcal{Q}} L^1(\Omega, \mathcal{B}, Q)$. The following statements hold.*

(a) If $\widetilde{\rho}$ is convex, then $\widetilde{\rho}^{\mathrm{WR}}$ and $\widetilde{\rho}^{\mathrm{MA}_2}$ are convex on L.

(b) If $\widetilde{\rho}$ satisfies comonotonic additivity, then $\widetilde{\rho}^{\mathrm{MA}_1}$ satisfies comonotonic additivity on L.

(c) If $\widetilde{\rho}$ satisfies translation invariance (positive homogeneity), then $\widetilde{\rho}^{\mathrm{WR}}$, $\widetilde{\rho}^{\mathrm{MA}_1}$ and $\widetilde{\rho}^{\mathrm{MA}_2}$ satisfy translation invariance (positive homogeneity) on L.

A direct result from Theorem 2 is that $\widetilde{\rho}^{\mathrm{WR}}$ and $\widetilde{\rho}^{\mathrm{MA}_2}$ are coherent whenever $\widetilde{\rho}$ is a coherent risk measure. Note that $\widetilde{\rho}^{\mathrm{WR}}$ may not be comonotonic additive even if $\widetilde{\rho}$ is. For instance, the mapping $\widetilde{\rho}^{\mathrm{WR}} : X \mapsto \max\{\mathbb{E}^{\mathbb{P}_1}[X], \mathbb{E}^{\mathbb{P}_2}[X]\}$ for $\mathbb{P}_1 \neq \mathbb{P}_2$ is not comonotonic additive in general.

Remark 3. Although all the risk measures considered in the paper are law-invariant with respect to $\mathbb{P}$, $\widetilde{\rho}^{\mathrm{WR}}$, $\widetilde{\rho}^{\mathrm{MA}_1}$ and $\widetilde{\rho}^{\mathrm{MA}_2}$ may not be law-invariant with respect to any one probability measure. Nevertheless, $\widetilde{\rho}^{\mathrm{WR}}$, $\widetilde{\rho}^{\mathrm{MA}_1}$ and $\widetilde{\rho}^{\mathrm{MA}_2}$ are all law-invariant with respect to the probability set $\mathcal{Q}$ according to the definitions of Delage et al. (2019) and Wang and Ziegel (2021).

5 MA approach in distributionally robust optimization

In this section, we consider optimization problems in which uncertainty is addressed by the WR and MA_2 approaches, i.e., $\preceq_2$ is chosen as the partial order. Let A be a set of possible actions, $f : A \times \mathbb{R}^d \to \mathbb{R}$ be a loss function, $\mathcal{F}$ be a set of cdfs on $\mathbb{R}^d$, and $\mathcal{F}_{\mathbf{a},f}$ is defined by (5). We consider the following two optimization problems

$$\min_{\mathbf{a}\in A} \ \rho^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f}) \quad \text{and} \quad \min_{\mathbf{a}\in A} \ \rho^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f}). \tag{16}$$

Recall the equivalency between the WR optimization and (6), i.e.,

$$\min_{\mathbf{a}\in A} \rho^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f}) = \min_{\mathbf{a}\in A} \sup_{F\in\mathcal{F}} \widetilde{\rho}^{F}(f(\mathbf{a},\mathbf{X})).$$

It is straightforward to see that the WR optimization is a convex problem if A, $\widetilde{\rho}$ and f (in its first argument) are convex because $\widetilde{\rho}$ is monotone and the inner problem is the supremum of a set of convex functionals. We demonstrate in the following result that the MA_2 optimization is convex under the same conditions.

Proposition 3. *Let A be a convex set. Suppose that a risk measure $\rho : \mathcal{M}_1 \to \mathbb{R}$ is convex, and $f : A \times \mathbb{R}^d \to \mathbb{R}$ is convex in its first argument. Then, the mapping $\mathbf{a} \mapsto \rho^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f})$ is convex. As a consequence, the MA_2 optimization problem in (16) is a convex one.*

We note that the convexity of both WR and MA_2 optimization problems may not always coincide; see Section E.2.

Next, we detail the application of the MA_2 approach to a broad class of coherent risk measures. We introduce an extra assumption before diving in.

Assumption 1. *For any $\mathbf{a} \in A$, the uncertainty set $\mathcal{F}_{\mathbf{a},f}$ is $\preceq_2$-bounded.*

Assumption 1 guarantees that $\bigvee_2 \mathcal{F}_{\mathbf{a},f}$ exists for any $\mathbf{a} \in A$. This assumption is mild. For example, it holds if we consider a common compact support of all involved cdfs.

A law-invariant coherent risk measure has a Kusuoka representation (Kusuoka (2001) and Shapiro (2013)), which can be approximated by the following form[11]

$$\rho(F) = \sup_{w\in\mathbb{W}} \sum_{j=1}^{n^w} p_j^w \mathrm{ES}_{\alpha_j^w}(F), \tag{17}$$

where $\mathbb{W}$ is a finite index set, and for each $w \in \mathbb{W}$, $n^w \in \mathbb{N}$, $(p_1^w, \dots, p_{n^w}^w) \in \Delta_{n^w}$ and $(\alpha_1^w, \dots, \alpha_{n^w}^w) \in [0,1]^{n^w}$. We formally show in Proposition EC.2 in Section E.1 that (17) can approximate any law-invariant coherent risk measure as the size of $\mathbb{W}$ goes infinity. For the risk measure in (17), we obtain a reformulation of the MA_2 optimization.

Theorem 3. *Suppose that Assumption 1 holds, and let ρ be given by (17). The MA_2 optimization*

[11] While our primary focus is on law-invariant coherent risk measures, we note that there is no extra difficulty to deal with a law invariant convex risk measure which has the form $\rho(F) = \sup_{w\in\mathbb{W}}\{\sum_{j=1}^{n^w} p_j^w \mathrm{ES}_{\alpha_j^w}(F) - \beta(w)\}$ where $\beta : \mathbb{W} \to \mathbb{R}$ is a penalty function.

problem in (16) can be reformulated as the following problem

$$\begin{aligned}
&\min_{\mathbf{a}\in A,\, x_j^w, h, h_j^w\in\mathbb{R}} \quad h \qquad (18)\\
&\qquad \text{s.t.} \quad \sum_{j=1}^{n^w} p_j^w h_j^w \leqslant h, \quad w\in\mathbb{W}\\
&\qquad\qquad x_j^w + \frac{1}{1-\alpha_j^w}\sup_{F\in\mathcal{F}}\mathbb{E}^F[(f(\mathbf{a},\mathbf{X})-x_j^w)_+] \leqslant h_j^w, \quad j\in[n^w],\ w\in\mathbb{W}.
\end{aligned}$$

Consider a finite uncertainty set $\mathcal{F}=\{F_1,\dots,F_n\}$, where $n\in\mathbb{N}$. Under this setting, we reformulate the MA_2 optimization, which is a direct consequence of Theorem 3. The two optimization problems in (16) can be respectively reformed as

$$\begin{aligned}
\mathrm{WR}: &\min_{\mathbf{a}\in A,\, x_{i,j}^w, h, h_{i,j}^w\in\mathbb{R}} \quad h \qquad (19)\\
&\qquad \text{s.t.} \quad \sum_{j=1}^{n^w} p_j^w h_{i,j}^w \leqslant h, \quad i\in[n],\ w\in\mathbb{W}\\
&\qquad\qquad x_{i,j}^w + \frac{1}{1-\alpha_j^w}\mathbb{E}^{F_i}[(f(\mathbf{a},\mathbf{X})-x_{i,j}^w)_+] \leqslant h_{i,j}^w, \quad i\in[n],\ j\in[n^w],\ w\in\mathbb{W};
\end{aligned}$$

$$\begin{aligned}
\mathrm{MA}_2: &\min_{\mathbf{a}\in A,\, x_j^w, h, h_j^w\in\mathbb{R}} \quad h \qquad (20)\\
&\qquad \text{s.t.} \quad \sum_{j=1}^{n^w} p_j^w h_j^w \leqslant h, \quad w\in\mathbb{W}\\
&\qquad\qquad x_j^w + \frac{1}{1-\alpha_j^w}\mathbb{E}^{F_i}[(f(\mathbf{a},\mathbf{X})-x_j^w)_+] \leqslant h_j^w, \quad i\in[n],\ j\in[n^w],\ w\in\mathbb{W}.
\end{aligned}$$

If $\rho=\mathrm{ES}_\alpha$ for $\alpha\in(0,1)$, then the two optimization problems in (16) can be reformed as

$$\begin{aligned}
\mathrm{WR}: &\min_{\mathbf{a}\in A,\, x_i, h\in\mathbb{R}} \quad h \quad \text{s.t.} \quad x_i + \frac{1}{1-\alpha}\mathbb{E}^{F_i}[(f(\mathbf{a},\mathbf{X})-x_i)_+] \leqslant h,\ i\in[n];\\
\mathrm{MA}_2: &\min_{\mathbf{a}\in A,\, x, h\in\mathbb{R}} \quad h \quad \text{s.t.} \quad x + \frac{1}{1-\alpha}\mathbb{E}^{F_i}[(f(\mathbf{a},\mathbf{X})-x)_+] \leqslant h,\ i\in[n].
\end{aligned}$$

A comparison between the computation time of (19) and (20) based on a specific example is given in Section E.2.

Remark 4. By Theorem 3 (or using Proposition 2), the MA_2 optimization retains the form (20) if we substitute the uncertainty set $\mathcal{F}=\{F_1,\dots,F_n\}$ with any set $\widetilde{\mathcal{F}}$, whose convex hull has extreme points $F_1,\dots,F_n$.

For other settings of Problem (18) that are solvable by convex program, see Section E.2. The two optimization problems in (16) are equivalent under a convex uncertainty set $\mathcal{F}$ when $\rho = \mathrm{ES}_\alpha$ for all $\alpha \in (0,1)$. This equivalence stems from the fact that $\mathcal{F}_{\mathbf{a},f}$ is convex for all $\mathbf{a} \in A$ which combining with Theorem 1 imply $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f}) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f})$ for all $\mathbf{a} \in A$. Many results in the literature rely on converting between $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f})$ and $\mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f})$; see e.g., Zhu and Fukushima (2009), Natarajan et al. (2010) and Cowell (2011).

6 Wasserstein and mean-variance uncertainty sets

In this section, we focus on three specific and popular uncertainty sets: (a) univariate Wasserstein uncertainty, (b) multivariate Wasserstein uncertainty, and (c) mean-variance uncertainty. We obtain explicit formulas for the robust models as well as the WR and MA robust risk evaluation. Furthermore, the portfolio selection problem will be explored based on these two robust approaches.

For results in this section and Section 8, we define a few classes of risk measures other than VaR and ES. The Range Value-at-Risk (RVaR), proposed by Cont et al. (2010), is defined as

$$\mathrm{RVaR}_{\alpha,\beta}(F) = \frac{1}{\beta-\alpha}\int_\alpha^\beta \mathrm{VaR}_s(F)\mathrm{d}s, \quad 0 \leqslant \alpha < \beta \leqslant 1.$$

Special and limiting cases of $\mathrm{RVaR}_{\alpha,\beta}$ include ES_α with $\beta = 1$ and VaR_β with $\alpha \uparrow \beta \in (0,1)$. If $\beta < 1$, then $\mathrm{RVaR}_{\alpha,\beta}$ is not $\preceq_2$-consistent by e.g., Wang et al. (2020, Theorem 3). The power-distorted (PD) risk measure (Wang (1995); Cherny and Madan (2009)) is defined as

$$\mathrm{PD}_k(F) = \int_0^1 ks^{k-1}\mathrm{VaR}_s(F)\mathrm{d}s, \quad k \geqslant 1.$$

The PD risk measure is coherent. The expectile, proposed by Newey and Powell (1987) and denoted by ex_α, is defined as the unique solution $t = \mathrm{ex}_\alpha(F_X) \in \mathbb{R}$ to the following equation,

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

where $x_- = \max\{-x, 0\}$ for $x \in \mathbb{R}$. The risk measure ex_α is coherent (and thus $\preceq_2$-consistent) if and only if $\alpha \in [1/2, 1)$; we will use this specification.

6.1 Uncertainty induced by the univariate Wasserstein metric

We first focus on an uncertainty set induced by the Wasserstein metric. Let $\mathcal{M}_p$ be the set of cdfs on $\mathbb{R}$ with finite pth moment and $F_0 \in \mathcal{M}_p$ be a pre-specified cdf used as benchmark. For

$p \geqslant 1$, the p-Wasserstein metric between F and F_0 is defined as

$$W_p(F, F_0) = \left(\int_0^1 |F^{-1}(s) - F_0^{-1}(s)|^p \mathrm{d}s \right)^{1/p}. \tag{21}$$

The corresponding uncertainty set is, for a parameter $\varepsilon \geqslant 0$,

$$\mathcal{F}_{p,\varepsilon}(F_0) = \{F \in \mathcal{M}_p : W_p(F, F_0) \leqslant \varepsilon\}, \tag{22}$$

which is a convex set. The parameter ε represents the magnitude of uncertainty. Denote by

$$F^1_{p,\varepsilon|F_0} = \bigvee_1 \mathcal{F}_{p,\varepsilon}(F_0) \quad \text{and} \quad F^2_{p,\varepsilon|F_0} = \bigvee_2 \mathcal{F}_{p,\varepsilon}(F_0)$$

the supremum of $\mathcal{F}_{p,\varepsilon}(F_0)$ with respect to $\preceq_1$ and $\preceq_2$, respectively. In the following result, we will identify an explicit form of the suprema $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ in terms of left quantile functions.

Theorem 4. *Suppose that $\varepsilon > 0$, $p \geqslant 1$ and $F_0 \in \mathcal{M}_p$.*

(a) The left quantile function of $F^1_{p,\varepsilon|F_0}$ is uniquely determined by

$$\int_\alpha^1 \left((F^1_{p,\varepsilon|F_0})^{-1}(\alpha) - F_0^{-1}(s) \right)_+^p \mathrm{d}s = \varepsilon^p, \quad \alpha \in (0,1). \tag{23}$$

(b) The set $\mathcal{F}_{1,\varepsilon}(F_0)$ is not $\preceq_2$-bounded. For $p > 1$, the left quantile function of $F^2_{p,\varepsilon|F_0}$ is given by

$$(F^2_{p,\varepsilon|F_0})^{-1}(\alpha) = F_0^{-1}(\alpha) + \left(1 - \frac{1}{p}\right)(1-\alpha)^{-1/p}\varepsilon, \quad \alpha \in (0,1). \tag{24}$$

Since the cdfs $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$, as well as their quantile functions, are obtained explicitly in Theorem 4, the robust risk values $\rho^{\mathrm{MA}_1}(\mathcal{F}_{p,\varepsilon}(F_0))$ and $\rho^{\mathrm{MA}_2}(\mathcal{F}_{p,\varepsilon}(F_0))$ can be computed in a straightforward manner. On the other hand, $\rho^{\mathrm{WR}}(\mathcal{F}_{p,\varepsilon}(F_0))$ is often difficult to compute if the risk measure is complicated, although there are some results in the literature that considered the WR approach for special choices of risk measures. Postek et al. (2016) presented combinations of risk measures and uncertainty sets that allow for computationally tractable reformulations.

As a feature of the robust model, both $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ are heavy-tailed even if the benchmark distribution F_0 is light-tailed. Heavy-tailed distributions are common for modeling financial data; see e.g., McNeil et al. (2015). Indeed, $(F^1_{p,\varepsilon|F_0})^{-1} \geqslant (F^2_{p,\varepsilon|F_0})^{-1}$, and $(F^2_{p,\varepsilon|F_0})^{-1}$ is the sum of the quantile F_0^{-1} and a Pareto quantile with tail index $p > 1$. Some other observations on the supremum distributions in Theorem 4 are made in Remark EC.2.

Noting that the Wasserstein uncertainty set $\mathcal{F}_{p,\varepsilon}(F_0)$ is convex, we have $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}_{p,\varepsilon}(F_0)) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}_{p,\varepsilon}(F_0))$ by Theorem 1. A simulation result in case of $p=2$, $\varepsilon=0.1$ and a standard normal benchmark distribution is reported in Section F.2.

6.2 Multivariate Wasserstein uncertainty

For $p \geqslant 1$ and $a \geqslant 1$, let $\mathcal{M}_p(\mathbb{R}^d)$ be the set of all cdfs on $\mathbb{R}^d$ with finite pth moment in each component. The p-Wasserstein metric on $\mathbb{R}^d$ between $F, G \in \mathcal{M}_p(\mathbb{R}^d)$ is defined as

$$W_{a,p}^d(F,G) = \inf_{F_{\mathbf{X}}=F, F_{\mathbf{Y}}=G} \left(\mathbb{E}[\|\mathbf{X}-\mathbf{Y}\|_a^p]\right)^{1/p},$$

where $\|\cdot\|_a$ is the L^a norm on $\mathbb{R}^d$; see e.g., Blanchet et al. (2022). If $d=1$, then $W_{a,p}^d$ is W_p in (21) where the infimum is attained by comonotonicity via the Fréchet-Hoeffding inequality. Define the Wasserstein uncertainty set for a benchmark distribution $F_0 \in \mathcal{M}_p(\mathbb{R}^d)$ as, similar to (22),

$$\mathcal{F}_{a,p,\varepsilon}^d(F_0) = \left\{F \in \mathcal{M}_p(\mathbb{R}^d) : W_{a,p}^d(F, F_0) \leqslant \varepsilon\right\}, \quad \varepsilon \geqslant 0. \tag{25}$$

We focus on a portfolio selection problem, i.e., the loss function is chosen as the linear function $f(\mathbf{w}, \mathbf{x}) = \mathbf{w}^\top \mathbf{x}$. The portfolio risk is $\rho(F_{\mathbf{w}^\top \mathbf{X}})$ for some weight vector $\mathbf{w} \in \mathbb{R}^d$ and risk vector $\mathbf{X}$ with unknown cdf in the multi-dimensional Wasserstein ball $\mathcal{F}_{a,p,\varepsilon}^d(F_0)$. The univariate uncertainty set for the cdf of $\mathbf{w}^\top \mathbf{X}$ is denoted by

$$\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_0) = \left\{F_{\mathbf{w}^\top \mathbf{X}} : F_{\mathbf{X}} \in \mathcal{F}_{a,p,\varepsilon}^d(F_0)\right\}, \quad F_0 \in \mathcal{M}_p(\mathbb{R}^d). \tag{26}$$

In the following theorem, we show that the problem of multivariate Wasserstein uncertainty can be conveniently converted to a univariate setting.

Theorem 5. *For $\varepsilon \geqslant 0$, $a \geqslant 1$ and $p \geqslant 1$, random vector $\mathbf{X}$ with $F_{\mathbf{X}} \in \mathcal{M}_p(\mathbb{R}^d)$ and $\mathbf{w} \in \mathbb{R}^d$, we have*

$$\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}}) = \mathcal{F}_{p,\varepsilon\|\mathbf{w}\|_b}(F_{\mathbf{w}^\top \mathbf{X}}),$$

where b satisfies $1/a + 1/b = 1$. As a consequence, for any $\rho : \mathcal{M}_1 \to \mathbb{R}$ and $i \in \{1,2\}$,

$$\rho^{\mathrm{WR}}(\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}})) = \rho^{\mathrm{WR}}(\mathcal{F}_{p,\varepsilon\|\mathbf{w}\|_b}(F_{\mathbf{w}^\top \mathbf{X}})) \quad \text{and} \quad \rho^{\mathrm{MA}_i}(\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}})) = \rho^{\mathrm{MA}_i}(\mathcal{F}_{p,\varepsilon\|\mathbf{w}\|_b}(F_{\mathbf{w}^\top \mathbf{X}})).$$

Intuitively, by Theorem 5, the multi-dimensional Wasserstein ball has the simple property of a usual Euclidean ball, that its affine projection is a lower-dimensional ball (this intuitive observation is not completely trivial because of the infimum in the Wasserstein metric). This result allows us to solve the MA robust portfolio optimization by applying Theorem 4.

We illustrate Theorem 5 with the setting of an elliptical benchmark distribution. An elliptical distribution with characteristic generator ψ is denoted by $\mathrm{E}(\boldsymbol{\mu}, \Sigma, \psi)$, which has normal and t-distributions as special cases; see McNeil et al. (2015, Chapter 6) for a precise definition. Let the benchmark distribution $F_0 = \mathrm{E}(\boldsymbol{\mu}, \Sigma, \psi)$ and denote by $F_\psi = \mathrm{E}(0, 1, \psi)$. Define a Pareto distribution G_p with $G_p^{-1}(\alpha) = (1-\alpha)^{-1/p}$ for $\alpha \in (0,1)$. By Theorems 4 and 5, it holds that

$$\min_{\mathbf{w}\in\mathcal{W}}: \quad \rho^{\mathrm{MA}_2}\left(\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_0)\right) = \widetilde{\rho}\left(\mathbf{w}^\top\boldsymbol{\mu} + \sqrt{\mathbf{w}^\top\Sigma\mathbf{w}}F_\psi^{-1}(U) + \left(1-\frac{1}{p}\right)\varepsilon\|\mathbf{w}\|_b G_p^{-1}(U)\right), \tag{27}$$

where U is a uniform random variable on $[0,1]$. The WR approach does not admit an explicit formula like (27), unless ρ is a coherent distortion risk measure; see Wozabal (2014) and Postek et al. (2016).

Consider a coherent distortion risk measure defined by $\rho_h(F) = \int_0^1 \mathrm{VaR}_s(F)\mathrm{d}h(s)$, where $h:[0,1]\to[0,1]$ is increasing and convex with $h(0) = 1 - h(1) = 0$. In this case, by applying Proposition 4 of Liu et al. (2022) and Theorems 4 and 5, we obtain the following reformulations

$$\min_{\mathbf{w}\in\mathcal{W}}: \quad \rho_h^{\mathrm{WR}}\left(\mathcal{F}_{\mathbf{w},p,\varepsilon}(F_0)\right) = \mathbf{w}^\top\boldsymbol{\mu} + \rho_h(F_\psi)\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} + \zeta(p,h)\varepsilon\|\mathbf{w}\|_b, \tag{28}$$

(this is also obtained by Wozabal (2014)) and

$$\min_{\mathbf{w}\in\mathcal{W}}: \quad \rho_h^{\mathrm{MA}_2}\left(\mathcal{F}_{\mathbf{w},p,\varepsilon}(F_0)\right) = \mathbf{w}^\top\boldsymbol{\mu} + \rho_h(F_\psi)\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} + \xi(p,h)\varepsilon\|\mathbf{w}\|_b, \tag{29}$$

where

$$\zeta(p,h) = \left(\int_0^1 (h'_+(s))^{p/(p-1)}\mathrm{d}s\right)^{(p-1)/p} \quad \text{and} \quad \xi(p,h) = \frac{p-1}{p}\int_0^1 (1-s)^{-1/p}\mathrm{d}h(s).$$

In particular, (28) and (29) are second-order conic program (SOCP) when $a = 2$; see e.g., Ben-Tal and Nemirovski (2001). Coherence of ρ (convexity of h) is essential for the WR formula in (28) because general formulas are not available for non-convex distortions under Wasserstein uncertainty. In contrast, the MA formula (29) holds for any distortion risk measures (even if they may not be $\preceq_2$-consistent) which directly follows from Theorems 4 and 5. Numerical and empirical results on the above approaches for robust portfolio selection are presented in Section 8.2.

6.3 Uncertainty induced by mean-variance information

Next, we pay attention to an uncertainty set defined by the first two moments, that is, for some $\mu \in \mathbb{R}$ and $\sigma > 0$, the set

$$\mathcal{F}_{\mu,\sigma} := \left\{F \in \mathcal{M}_2 : \mathfrak{m}(F) = \mu \text{ and } \operatorname{var}(F) = \sigma^2\right\}, \tag{30}$$

where $\mathfrak{m}(F)$ and $\operatorname{var}(F)$ represent the mean and the variance of F, respectively. The two equalities in (30) can be safely replaced by inequalities $\mathfrak{m}(F) \leqslant \mu$ and $\operatorname{var}(F) \leqslant \sigma^2$ in the problems we consider, and we omit the formulation with inequalities. The WR robust risk value for different risk measures based on this uncertainty set $\mathcal{F}_{\mu,\sigma}$ has been extensively studied in literature, see e.g., El Ghaoui et al. (2003), Zhu and Fukushima (2009), Natarajan et al. (2010), Cowell (2011), Li (2018) and the references therein.

For the MA approach, we will identify the supremum of $\mathcal{F}_{\mu,\sigma}$ with respect to $\preceq_1$ and $\preceq_2$. Theorem 1 of El Ghaoui et al. (2003) and Corollary 1.1 of Jagannathan (1977) (see also Müller and Stoyan (2002, Theorem 1.10.7)) yield

$$\sup_{F \in \mathcal{F}_{\mu,\sigma}} \mathrm{VaR}_\alpha(F) = \mu + \sigma\sqrt{\frac{\alpha}{1-\alpha}}, \quad \alpha \in (0,1)$$

and

$$\sup_{F \in \mathcal{F}_{\mu,\sigma}} \pi_F(x) = \frac{1}{2}\left(\mu - x + \sqrt{(x-\mu)^2 + \sigma^2}\right), \quad x \in \mathbb{R}.$$

Denote by $F^1_{\mu,\sigma} = \bigvee_1 \mathcal{F}_{\mu,\sigma}$ and $F^2_{\mu,\sigma} = \bigvee_2 \mathcal{F}_{\mu,\sigma}$ the supremum of $\mathcal{F}_{\mu,\sigma}$ with respect to $\preceq_1$ and $\preceq_2$, respectively. Using Proposition 1 and above two equations, we immediately get the explicit expressions of $F^1_{\mu,\sigma}$ and $F^2_{\mu,\sigma}$.

Proposition 4. *Suppose that $\mu \in \mathbb{R}$ and $\sigma > 0$. We have*

$$F^1_{\mu,\sigma}(x) = \frac{(x-\mu)^2}{\sigma^2 + (x-\mu)^2}, \quad x \geqslant \mu; \tag{31}$$

$$F^2_{\mu,\sigma}(x) = \frac{1}{2}\left(1 + \frac{x-\mu}{\sqrt{(x-\mu)^2 + \sigma^2}}\right), \quad x \in \mathbb{R}. \tag{32}$$

We note that both $F^1_{\mu,\sigma}$ and $F^2_{\mu,\sigma}$ are in $\mathcal{M}_1$, so they are ready for implementation with any risk measures or preferences well-defined on $\mathcal{M}_1$; however, none of $F^1_{\mu,\sigma}$ and $F^2_{\mu,\sigma}$ is in $\mathcal{M}_2$. Most risk measures in practice, including ES and VaR and the other examples in this section, are well-defined and finite on $\mathcal{M}_1$.

Table 1: WR and MA under uncertainty induced by $\mathcal{F}_{0,1}$

ρ	ρ^{WR}	ρ^{MA_1}	ρ^{MA_2}
ES_α	$\sqrt{\frac{\alpha}{1-\alpha}}$	$\frac{1}{1-\alpha}\int_\alpha^1 \sqrt{\frac{s}{1-s}}\mathrm{d}s$	$\sqrt{\frac{\alpha}{1-\alpha}}$
$\mathrm{RVaR}_{\alpha,\beta}$	$\sqrt{\frac{\alpha}{1-\alpha}}$	$\frac{1}{\beta-\alpha}\int_\alpha^\beta \sqrt{\frac{s}{1-s}}\mathrm{d}s$	$\frac{1}{\beta-\alpha}\int_\alpha^\beta \frac{s-1/2}{\sqrt{s(1-s)}}\mathrm{d}s$
VaR_α	$\sqrt{\frac{\alpha}{1-\alpha}}$	$\sqrt{\frac{\alpha}{1-\alpha}}$	$\frac{\alpha-1/2}{\sqrt{\alpha(1-\alpha)}}$
PD_k	$\frac{k-1}{\sqrt{2k-1}}$	$\frac{\sqrt{\pi}\Gamma(k+1/2)}{\Gamma(k)}$	$\frac{\sqrt{\pi}(k-1)}{2k-1}\frac{\Gamma(k+1/2)}{\Gamma(k)}$
ex_α	$\frac{\alpha-1/2}{\sqrt{\alpha(1-\alpha)}}$	$\mathrm{ex}_\alpha(F^1_{0,1})$	$\frac{\alpha-1/2}{\sqrt{\alpha(1-\alpha)}}$

Note: Γ is the gamma function; $\mathrm{ex}_\alpha(\mathcal{F}^1_{0,1})$ can be numerically computed but it does not admit an explicit formula.

By Proposition 4, for a risk measure that is $\preceq_1$-consistent or $\preceq_2$-consistent, the MA robust risk value for the uncertainty set $\mathcal{F}_{\mu,\sigma}$ can be directly obtained by calculating the risk measure of $F^1_{\mu,\sigma}$ or $F^2_{\mu,\sigma}$. To compute the WR robust risk value, for a coherent risk measure ρ, Li (2018) gives the explicit expression of $\rho^{\mathrm{WR}}(\mathcal{F}_{\mu,\sigma})$ based on the Kusuoka representation. In addition, noting that $\mathcal{F}_{\mu,\sigma}$ is a convex set, if ρ is an ES, then $\rho^{\mathrm{WR}}(\mathcal{F}_{\mu,\sigma}) = \rho^{\mathrm{MA}_2}(\mathcal{F}_{\mu,\sigma}) = \rho(F^2_{\mu,\sigma})$. If ρ is a VaR, then $\rho^{\mathrm{WR}}(\mathcal{F}_{\mu,\sigma}) = \rho^{\mathrm{MA}_1}(\mathcal{F}_{\mu,\sigma}) = \rho(F^1_{\mu,\sigma})$. The explicit WR and MA robust risk values for ES_α, $\mathrm{RVaR}_{\alpha,\beta}$, the power-distorted risk measure and the expectile are given in Table 1,[12] and a few figures on their numerical values are reported in Section F.2. Since those risk measures satisfy translation invariance and positive homogeneity, it suffices to consider the case $(\mu,\sigma) = (0,1)$.

Similar to Section 6.2, we apply the MA approach with mean-variance uncertainty to robust portfolio selection. The portfolio risk is $\rho(F_{\mathbf{w}^\top\mathbf{X}})$ for some portfolio weight vector $\mathbf{w} \in \mathbb{R}^d$ and risk vector $\mathbf{X}$ with unknown cdf in the uncertainty set with given first two moments, which can be formulated as, for a feasible set $\mathcal{W}$ of $\mathbf{w}$,

$$\min_{\mathbf{w}\in\mathcal{W}} \; \rho^{\mathrm{MA}}\left(\mathcal{F}_{\mathbf{w},\boldsymbol{\mu},\Sigma}\right), \quad \text{where } \mathcal{F}_{\mathbf{w},\boldsymbol{\mu},\Sigma} = \{F_{\mathbf{w}^\top\mathbf{X}} : \mathbb{E}[\mathbf{X}] = \boldsymbol{\mu},\ \mathrm{Cov}(\mathbf{X}) = \Sigma\}, \tag{33}$$

where $\mathbb{E}[\mathbf{X}]$ and $\mathrm{Cov}(\mathbf{X})$ represents the mean vector and the covariance of $\mathbf{X}$. Applying the general projection property in Popescu (2007) (see also Cowell (2011, Lemma 2.2)), the two sets $\mathcal{F}_{\mathbf{w},\boldsymbol{\mu},\Sigma}$

[12] To obtain these formulas, we use the following results. Li et al. (2018) showed that $\mathrm{RVaR}^{\mathrm{WR}}_{\alpha,\beta}(\mathcal{F}_{\mu,\sigma}) = \mathrm{ES}^{\mathrm{WR}}_\alpha(\mathcal{F}_{\mu,\sigma})$ for all $\beta \in (\alpha,1)$. The value of PD_k via the WR approach can be directly derived by Li (2018, Theorem 2). An expectile can be represented as the supremum of convex combinations of ES and expectation; see Bellini et al. (2014, Proposition 9). By Theorem 6 and noting that all elements in $\mathcal{F}_{\mu,\sigma}$ have the same expectation, we obtain $\mathrm{ex}^{\mathrm{WR}}_\alpha(\mathcal{F}_{\mu,\sigma}) = \mathrm{ex}^{\mathrm{MA}_2}_\alpha(\mathcal{F}_{\mu,\sigma})$.

and $\mathcal{F}_{\mathbf{w}^\top \boldsymbol{\mu}, \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}$ are identical. Hence, (33) is equivalent to

$$\min_{\mathbf{w}\in\mathcal{W}} \quad \rho^{\mathrm{MA}}\left(\mathcal{F}_{\mathbf{w}^\top \boldsymbol{\mu}, \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}\right).$$

In case of MA_1 or MA_2, this leads to

$$\min_{\mathbf{w}\in\mathcal{W}}: \quad \rho^{\mathrm{MA}_i}\left(\mathcal{F}_{\mathbf{w}^\top \boldsymbol{\mu}, \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}\right) = \widetilde{\rho}^{F^i_{0,1}}\left(\mathbf{w}^\top \boldsymbol{\mu} + \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}} X\right), \tag{34}$$

where $F^i_{0,1}$ is given by Proposition 4 in explicit form for $i = 1, 2$. In particular, if ρ satisfies translation invariance and positive homogeneity, Problem (34) leads to the following convenient formulation of SOCP, for $i = 1, 2$,

$$\min_{\mathbf{w}\in\mathcal{W}}: \quad \rho^{\mathrm{MA}_i}\left(\mathcal{F}_{\mathbf{w}^\top \boldsymbol{\mu}, \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}\right) = \left\{\mathbf{w}^\top \boldsymbol{\mu} + \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}\ \rho\left(F^i_{0,1}\right)\right\}.$$

7 Characterization of risk measures by equivalence in MA

In this section, we aim to characterize the risk measures under which the WR and MA approaches are equivalent, that is,

$$\rho\left(\bigvee \mathcal{F}\right) = \sup_{F\in\mathcal{F}} \rho(F), \tag{35}$$

for all $\mathcal{F} \in \mathcal{S}$, where $\mathcal{S}$ is a collection of subsets of $\mathcal{M}$. We are interested in the case that $\mathcal{S}$ is the collection of all convex polytopes which is defined by

$$\mathcal{F} = \mathrm{conv}(F_1, \dots, F_n) = \left\{\sum_{i=1}^n \lambda_i F_i : \boldsymbol{\lambda} \in \Delta_n\right\},$$

where $F_1, \dots, F_n$ are finitely many cdfs. The corresponding property is called *equivalence in model aggregation for convex polytopes* (cEMA); that is, (35) holds for all convex polytopes $\mathcal{F} \subseteq \mathcal{M}$.

> $\preceq$-*cEMA*: Let $(\mathcal{M}, \preceq)$ be an ordered set. A mapping $\rho : \mathcal{M} \to \mathbb{R}$ satisfies $\preceq$-cEMA if $\rho\left(\bigvee \mathcal{F}\right) = \sup_{F\in\mathcal{F}} \rho(F)$ holds for any nonempty convex polytope $\mathcal{F} \subseteq \mathcal{M}$.

All results in this section remain valid if convex polytopes in the above definition are replaced by convex sets bounded from above, and such a property is stronger than cEMA.[13]

[13] Recall that characterization results are generally stronger if imposed properties are weaker, so we aim for a weaker formulation of the properties.

Our main focus is the partial orders $\preceq_1$ and $\preceq_2$. By Proposition 2, $\preceq_i$-cEMA is equivalent to

$$\rho\left(\bigvee_i \{F_1, \dots, F_n\}\right) = \sup\left\{\rho\left(\sum_{i=1}^n \lambda_i F_i\right) : \boldsymbol{\lambda} \in \Delta_n\right\}, \quad i = 1, 2, \tag{36}$$

for all $F_1, \dots, F_n \in \mathcal{M}$. By (36), $\preceq_i$-cEMA is stronger than $\preceq_i$-consistency since for any $F \preceq_i G$, we have $\rho(G) = \rho\left(\bigvee_i \{F, G\}\right) = \sup_{\lambda \in [0,1]} \rho(\lambda F + (1-\lambda)G) \geqslant \rho(F)$ under $\preceq_i$-cEMA.

By Theorem 1, VaR satisfies $\preceq_1$-cEMA, and ES satisfies $\preceq_2$-cEMA. The more challenging question is in the opposite direction: Are VaR and ES the unique classes of risk measures, with some standard properties, that satisfies $\preceq_1$-cEMA and $\preceq_2$-cEMA, respectively? This question is particularly important given the special roles of VaR and ES in banking practice. We obtain two main results: With the additional standard properties of translation invariance, positive homogeneity, and lower semicontinuity, $\preceq_1$-cEMA characterizes VaR, and $\preceq_2$-cEMA characterizes ES. As far as we are aware, this is the first time that VaR and ES are axiomatized with parallel properties.

Theorem 6. *For a mapping $\rho : \mathcal{M}_1 \to \mathbb{R}$,*

(a) it satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_1$-cEMA if and only if $\rho = \mathrm{VaR}_\alpha$ for some $\alpha \in (0, 1)$;

(b) it satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_2$-cEMA if and only if $\rho = \mathrm{ES}_\alpha$ for some $\alpha \in (0, 1)$.

The special case of $\mathrm{ES}_0 = \mathfrak{m}$ is excluded from Theorem 6, as it satisfies $\preceq_2$-cEMA (by Theorem 1) but not lower semicontinuity. Theorem 6 states that $\preceq_1$-cEMA and $\preceq_2$-cEMA can identify VaR and ES, respectively. In contrast to VaR which satisfies (35) for any $\mathcal{F}$ bounded from above (Theorem 1), ES fails to satisfy (35) for non-convex set $\mathcal{F}$ (Example 1).

Remark 5. There are a few sets of axioms which characterize VaR, each with the additional help of some standard properties such as continuity, monotonicity, translation invariance or positive homogeneity. In Chambers (2009), the main axiom for VaR is ordinal covariance, an invariance property under some risk transforms. In Kou and Peng (2016), the main axioms for VaR are elicitability and comonotonic additivity. In He and Peng (2018), the main axiom for VaR is surplus-invariance of the acceptance set. In Liu and Wang (2021), the main axioms are tail relevance and elicitability. In Theorem 6, the new axiom of $\preceq_1$-cEMA leads to a characterization of VaR, and this new axiom standalone does not imply any axioms mentioned above.

Remark 6. ES is recently axiomatized by Wang and Zitikis (2021) in the context of portfolio capital requirement. Their key axiom is called no reward for concentration (NRC) which intuitively means that a concentrated portfolio does not receive a diversification benefit. Han et al. (2024), who also considered concentrated portfolio, obtain another characterization of ES by relaxing NRC. Another characterization result on ES is obtained by Embrechts et al. (2021) based on elicitability and Bayes risk. In contrast, our characterization result does not involve the consideration of elicitability or portfolio risk aggregation. Therefore, the interpretation of Theorem 6 is quite different from results in the literature and can be applied to robust modeling outside of a financial or statistical context.

Remark 7. Equivalence in model aggregation has some similarity to max-stability studied by Kupper and Zapata (2021), which is defined on the set of random variables with the natural order, i.e., $X \preceq Y$ if and only if $X \leqslant Y$ pointwisely. This leads to completely different interpretations and mathematics.

8 Numerical results for financial data

In this section, we report some numerical experiments based on real financial data to show the performance of the MA approach. We select 20 stocks and collect their historical loss data from Yahoo! Finance.[14] We use the period of January 1, 2019, to August 1, 2021, with a total of 649 observations of the daily losses of the 20 stocks.

We shall conduct two sets of numerical experiments. First, in Section 8.1, we present the robust distributions based on the MA approach when the uncertainty set consists of finite cdfs generated from the historical data, and compare the robust risk values with the WR ones. This analysis is based on data of single asset, and we only report results on AAPL for a simple illustration. Second, in Section 8.2, we consider the application of the MA approach with Wasserstein and mean-variance uncertainty as in Section 6, and data of all 20 stocks will be used.

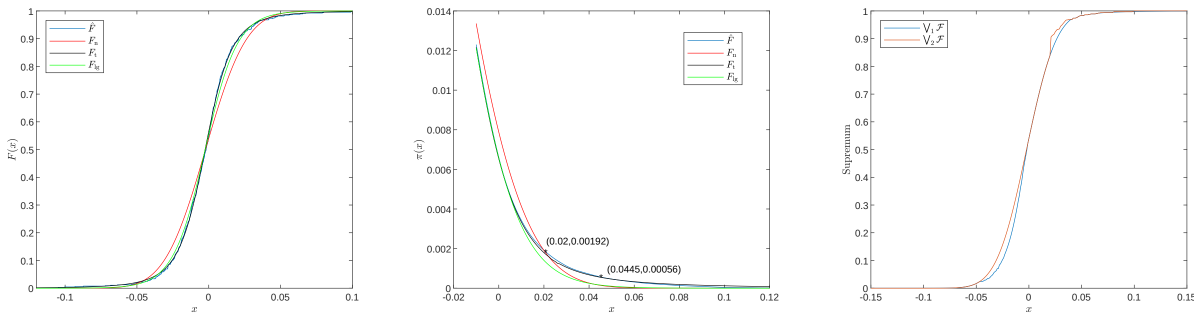
8.1 Performance of MA with finite uncertainty set

We examine the MA approach for the uncertainty set that consists of the cdfs generated by the real portfolio data AAPL. We use `Matlab` to fit the data with normal, t- and logistic distributions that will be denoted by F_{n}, F_{t} and F_{lg}, respectively, and the empirical cdf is denoted by $\hat{F}$. Let $\mathcal{F}$ be the uncertainty set that consists of these four cdfs, i.e., $\mathcal{F} = \{\hat{F}, F_{\mathrm{n}}, F_{\mathrm{t}}, F_{\mathrm{lg}}\}$.

Figure 1 (top panels) shows the cdfs and integrated survival functions defined by (10) of the cdfs in $\mathcal{F}$. Noting that $\bigvee_1 \mathcal{F}(x) = \inf\{\hat{F}(x), F_{\mathrm{n}}(x), F_{\mathrm{t}}(x), F_{\mathrm{lg}}(x)\}$ for $x \in \mathbb{R}$, the supremum $\bigvee_1 \mathcal{F}$ can be roughly divided into four parts. By Proposition 1, $\bigvee_2 F = F^* \in \mathcal{F}$ on (a, b) if F^* has

[14]They are AAPL, MSFT, GOOGL, AMZN, ADBE, NFLX, AMD, V, JNJ, COST, WMT, PG, MA, UNH, DIS, HD, INTC, PYPL, GS, IBM.

Figure 1: Left: cdf; Middle: Integrated survival function; Right: Suprema of $\mathcal{F}$



the largest value of integrated survival function on (a, b). Hence, the figure of integrated survival functions illustrates $\bigvee_2 \mathcal{F}$ can be divided into three parts: $\bigvee_2 \mathcal{F} = F_{\rm n}$ on $(-\infty, 0.02)$; $\bigvee_2 \mathcal{F} = \hat{F}$ on $[0.02, 0.0445)$; $\bigvee_2 \mathcal{F} = F_{\rm t}$ on $[0.0445, \infty)$. The curves of $\bigvee_1 \mathcal{F}$ and $\bigvee_2 \mathcal{F}$ are given in Figure 1 (bottom panel) from which we can see that $\bigvee_2 \mathcal{F} \preceq_1 \bigvee_1 \mathcal{F}$. Moreover, $\bigvee_2 \mathcal{F}$ has a jump at 0.02 which can be explained by the difference between left and right derivatives of the integrated survival function of $\bigvee_2 \mathcal{F}$ at 0.02.

In the following, we compare the MA_1 and MA_2 robust risk values and the WR ones with the uncertainty set $\mathcal{F}$. The risk measures are RVaR or ES. In the case of $\mathrm{RVaR}_{\alpha,\beta}$, we set $\alpha = 0.95$ and let β vary in $[0.95, 1]$. In the case of ES_α, α varies in $[0.9, 0.99]$.

Figure 2 shows the value of $\mathrm{RVaR}_{\alpha,\beta}$ of the cdfs in $\mathcal{F}$, and $\mathrm{RVaR}_{\alpha,\beta}$ based on the MA and WR approaches, and Figure 3 presents the results of ES. From both figures we can see that the MA robust risk value is larger than the WR one. Moreover, from Figure 2, one can find that these two robust approaches have identical performance for $\beta \in [0.95, 0.9685]$. This is because the quantile function of $F_{\rm n}$ dominates other elements in $\mathcal{F}$ on $[0.95, 0.9685]$ which implies that $\mathrm{RVaR}^{\mathrm{MA}_1}_{0.95,\beta}(\mathcal{F}) = \mathrm{RVaR}^{\mathrm{WR}}_{0.95,\beta}(\mathcal{F}) = \mathrm{RVaR}_{0.95,\beta}(F_{\rm n})$ for $\beta \in [0.95, 0.9685]$. From Figure 3, we find that $\mathrm{ES}_\alpha(F_{\rm n})$ and $\mathrm{ES}_\alpha(F_{\rm lg})$ are always smaller than $\mathrm{ES}_\alpha(\hat{F})$ and $\mathrm{ES}_\alpha(F_{\rm t})$ for $\alpha \in [0.9, 0.99]$. The reason is that financial market loss data are heavy-tailed empirically (see e.g., McNeil et al. (2015)), and ES with high level α focuses on the tail loss. In addition, the curve of $\mathrm{ES}^{\mathrm{MA}}$ always lies above the curve of $\mathrm{ES}^{\mathrm{WR}}$, which implies that the MA approach is more conservative.

8.2 MA approach in robust portfolio selection

In this section, we consider the application of the MA approach with $\preceq_2$ in the setting of portfolio selection in Section 6. The MA approach will be contrasted to the WR approach, and the

Figure 2: RVaR for individual models, via WR and via MA (α=0.95)

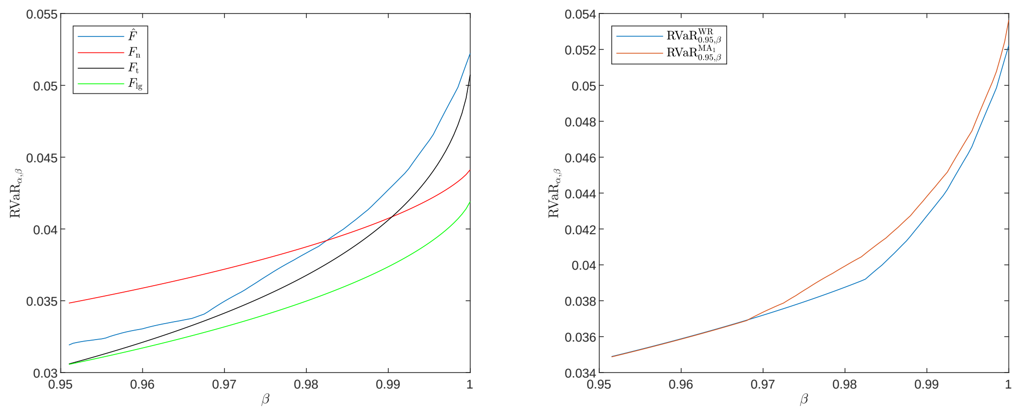
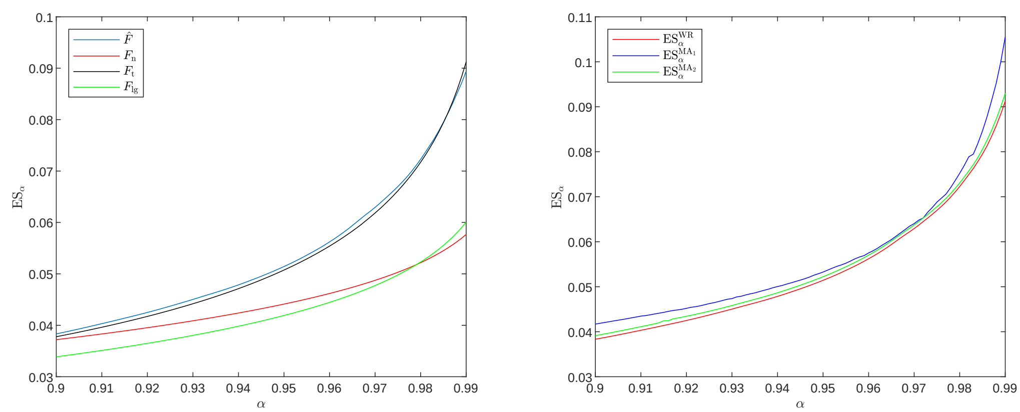


Figure 3: ES for individual models, via WR and via MA



standard sample average approximation (SAA) approach, which evaluates risks using the empirical distribution; see e.g., Shapiro et al. (2021). We construct a portfolio from the 20 stocks mentioned in the beginning of this section, whose daily losses are denoted by X_1 (AAPL), X_2 (MSFT), ..., X_{20} (IBM). Mean, variance and correlation matrix of the return rate of the 20 stocks are given in Appendix H. The wealth invested in the asset X_i is denoted by w_i for $i \in [20]$. Thus, the total loss from the investment of these 20 stocks is $\mathbf{w}^\top \mathbf{X}$, where $\mathbf{w} = (w_1, \dots, w_{20})$ and $\mathbf{X} = (X_1, \dots, X_{20})$. The feasible region of $\mathbf{w}$ is the standard simplex Δ_{20}.

We consider the setting in Section 6 where uncertainty is modeled by a multi-dimensional Wasserstein ball. For the choice of the risk measure ρ, we will work with PD_k defined in Section 6 to measure the portfolio risk. There are a few reasons for this choice. First, PD_k is $\preceq_2$-consistent (which also implies $\preceq_1$-consistency). Second, the WR and MA_2 approaches are similar in the portfolio optimization problem under the Wasserstein or the mean-variance uncertainty if the risk measure is selected as ES or expectile, so we move away from these two choices. Third, the portfolio optimization problem of PD_k leads to a convenient formulation of SOCP under the Wasserstein or the mean-variance uncertainty as in Section 6.

As in many classic settings of portfolio selection, e.g., the classic framework of Markowitz (1952), we assume that the investor has a target level of expected annualized return rate and minimizes the risk. That is, with the constraint $\mathbb{E}[\mathbf{w}^\top \mathbf{X}] \leqslant -r_0/m$ where r_0 is the expected annualized return rate and $m = 250$, the investor minimizes $\rho(F_{\mathbf{w}^\top \mathbf{X}})$.

We set the parameter $a = p = 2$ in the Wasserstein uncertainty ball $\mathcal{F}^d_{a,p,\varepsilon}(F_0)$, and use a multivariate t-benchmark distribution F_0 fitted to the data. The case of a normal benchmark distribution, which has a lighter tail, is reported in Appendix H. For the whole-period data, the fitted t-distribution has $\nu = 3.994$ degrees of freedom. The choice of a t-distribution is by no means restrictive, and we will consider the case of normal distribution which has a lighter tail in Appendix. We apply the WR and the $\preceq_2$-MA approaches, and the corresponding portfolio optimization problems are converted to SOCPs which can be computed efficiently. By (28) and (29) in Section 6.2, the optimization problems via the WR and the MA approaches are, respectively,

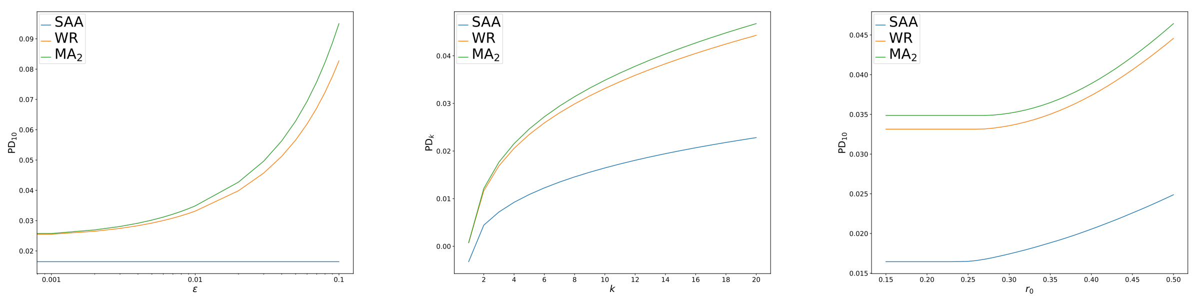
$$\min_{\mathbf{w}\in\Delta_{20}} : \quad \rho^{\mathrm{WR}}\left(\mathcal{F}_{\mathbf{w},2,2,\varepsilon}(F_0)\right) = \mathbf{w}^\top\boldsymbol{\mu} + \mathrm{PD}_k(F_\nu)\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} + \zeta_k\varepsilon\sqrt{\mathbf{w}^\top\mathbf{w}} \quad \text{s.t. } \mathbf{w}^\top\boldsymbol{\mu} \leqslant -r_0/m, \tag{37}$$

and

$$\min_{\mathbf{w}\in\Delta_{20}} : \quad \rho^{\mathrm{MA}_2}\left(\mathcal{F}_{\mathbf{w},2,2,\varepsilon}(F_0)\right) = \mathbf{w}^\top\boldsymbol{\mu} + \mathrm{PD}_k(F_\nu)\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} + \xi_k\varepsilon\sqrt{\mathbf{w}^\top\mathbf{w}} \quad \text{s.t. } \mathbf{w}^\top\boldsymbol{\mu} \leqslant -r_0/m, \tag{38}$$

where $\zeta_k = k/\sqrt{2k-1}$, $\xi_k = (\sqrt{\pi}\Gamma(k+1))/(2\Gamma(k+1/2))$, F_ν is the unit variance t-distribution with the tail parameter ν, and $\boldsymbol{\mu}$ and Σ are the mean and the covariance of the fitted F_0, i.e., the mean vector and the covariance of the whole-period data. The SAA approach optimizes the portfolio according to the empirical cdf of the asset losses. Figure 4 presents the optimized risk value of SAA approach and the optimized robust risk values under Wasserstein uncertainty with the WR and MA approaches for different values of ε, r_0 and k using the whole-period data. In the left panel, the MA robust risk value is larger than the WR one, and both are generally larger than that of SSA. This is consistent with our intuition as MA is more conservative than WR, and SAA is not a conservative method. In the middle and the right panels we set $\varepsilon = 0.01$ and let k and r

Figure 4: The optimized robust values of PD_k under Wasserstein uncertainty using the whole-period data. Left: $r_0 = 0.2$, $k = 10$, $\varepsilon \in [0, 0.1]$; Middle: $r_0 = 0.2$, $\varepsilon = 0.01$, $k \in [1, 20]$; Right: $k = 10$, $\varepsilon = 0.01$, $r_0 \in [0.15, 0.5]$



vary. In practice, the parameter ε should not be too small; one may tune ε so that the empirical cdf remains in the Wasserstein ball.

We choose slightly more than half of the period (350 trading days) for the initial training, and optimize the portfolio weights in each day with a rolling window. That is, on each trading day starting from day 351 (roughly June 2020), the preceding 350 trading days are used to fit the benchmark distribution, and compute the optimal portfolio weights. Note that the parameter k reflects the degree of risk aversion of the decision maker, that is, a larger value of k indicates a more risk-averse decision maker, and thus a larger corresponding risk measure. In this experiment, we choose $k = 2$ and 20, and the decision maker with $k = 20$ is more risk-averse than the one with $k = 2$. Figure 5 depicts the performance of SAA approach, the mean-variance model of Markowitz (1952) (minimizing the variance of $\mathbf{w}^\top \mathbf{X}$ subject to $\mathbb{E}[\mathbf{w}^\top \mathbf{X}] \leqslant -r_0/m$), and the MA and WR approaches under Wasserstein uncertainty over the remaining 300 trading days with $r_0 = 0.2$ and $\varepsilon = 0.01$, and we set $k = 2$ (left) or $k = 20$ (right). Table 2 presents realized annualized return rates and Sharpe ratios of all methods. In all results, the MA and WR approaches, being robust methods, perform similarly. MA and WR generally outperform the SAA and the Markowitz model, especially after the first 150 trading days. Intuitively, this means that, during the period from Jan to Aug 2021, robust investment strategies likely outperform non-robust strategies. The similar performance of the MA and WR approaches under Wasserstein uncertainty is not a coincidence due to the similarity of problems (37) and (38) by noting that ε is small.

Table 3 presents the nominal transaction cost for different strategies by using the average weight change $\sum_{t=1}^{T} \|\mathbf{w}_{t+1} - \mathbf{w}_t\|_1 / T$ where $\mathbf{w}_t$ is the weight used on day $t \in [T]$ by each strategy (see Olivares-Nadal and DeMiguel (2018)). The MA and WR approaches based on Wasserstein

Figure 5: Wealth evolution for different portfolio strategies from May 2020 to Aug 2021 ($\varepsilon = 0.01$, $r_0 = 0.2$). Left: $k = 2$; Right: $k = 20$

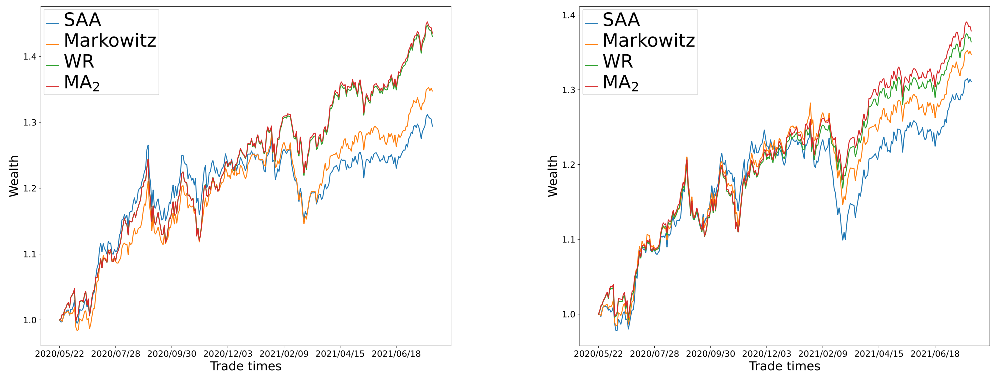


Table 2: Annualized return (AR), annualized volatility (AV) and Sharpe ratio (SR) for different strategies from May 2020 to Aug 2021; $r_0 = 0.2$ and the risk-free rate is 0.165% (1-year US Treasury yield on May 22, 2020)

	AR (%)		AV (%)		SR (%)	
Approach	$k = 2$	$k = 20$	$k = 2$	$k = 20$	$k = 2$	$k = 20$
SAA	25.42	26.72	14.82	14.35	170.4	185.0
Markowitz	29.50	29.50	**13.54**	13.54	216.6	216.6
WR	32.75	30.79	14.25	**13.30**	228.7	230.3
MA_2	**33.77**	**32.01**	14.48	13.56	**232.0**	**234.9**

uncertainty have similar transaction costs, which are smaller than the other methods in most cases. A similar analysis using the mean-variance uncertainty is reported in Appendix H.

9 Concluding remarks and discussions

The MA approach for robust risk evaluation is proposed. Below, we summarize some the advantages of the MA approach, which are illustrated and discussed through several technical results.

1. The MA approach is natural to interpret, and it is motivated by the need for a robust distributional model. The WR approach is also natural to interpret, but the focus is on the risk

Table 3: Nominal transaction cost $\sum_{t=1}^{T} \|\mathbf{w}_{t+1} - \mathbf{w}_t\|_1 / T$ with $\varepsilon = 0.01$ and $T = 299$

	$r_0 = 0.1$		$r_0 = 0.2$		$r_0 = 0.3$	
Approach	$k = 2$	$k = 20$	$k = 2$	$k = 20$	$k = 2$	$k = 20$
SAA	0.0549	0.0071	0.0871	0.0121	0.0969	0.0833
Markowitz	**0.0102**	0.0102	0.0110	0.0110	0.0746	0.0746
WR	0.0127	**0.0032**	0.0127	**0.0033**	0.0271	0.0382
MA_2	0.0114	0.0035	**0.0105**	0.0035	**0.0239**	**0.0348**

value instead of the risk model. Different from the WR approach, the MA approach is built on stochastic orders and lattice theory. The robust model produced by the MA approach can be readily applied to different risk evaluation procedures and decision problems (Section 3).

2. The MA robust risk value is straightforward to compute (Section 4.1). In some settings of uncertainty, the MA approach leads to explicit formulas for the robust model (Section 6). In particular, it can handle Wasserstein uncertainty in portfolio selection, based on a new dimension reduction result on Wasserstein balls (Theorem 5).

3. The MA approach admits reformulations in distributionally robust optimization similar to the WR approach, and it leads to a convex program when the loss function and the risk measure are convex (Section 5).

4. The MA approach gives rise to the useful property of cEMA which characterizes VaR and ES (Section 7). These results reveal a profound connection of the popular regulatory risk measures to robust risk evaluation methods, and highlight the special roles of VaR and ES among all risk measures, which is in itself a highly active research topic in risk management.

The MA approach requires a stochastic order to be specified. For an interpretation of prudent risk evaluation as in (3), the risk measure of interest should be consistent with this stochastic order. We recommend, in most applications, using $\preceq_2$ in an MA approach as the default option, for its nice interpretation in decision theory (strong risk aversion) and mathematical properties as developed in this paper.

We have focused on studying the MA and WR approaches together with risk measures throughout the paper. Both approaches can be easily applied to other objectives other than risk measures, such as expected utility functions, rank-dependent expected utilities, or other behaviour decision criteria. Some decision criteria may work better with notions of stochastic dominance other than FSD and SSD, and they may include considerations of model uncertainty by design; see e.g., Hansen and Sargent (2001), Maccheroni et al. (2006) and Cerreia-Vioglio et al. (2021).

Our theory is built on model spaces of univariate cdfs on $\mathbb{R}$ for the following reasons. First, classic risk measures, especially the ones used in regulatory practice such as VaR and ES, are defined on one-dimensional cdfs representing potential (portfolio) losses; second, commonly used stochastic orders, the key tool to build robust model aggregation in this paper, are usually defined on one-dimensional cdfs and they are naturally interpretable in this setting; third, many problems that are multivariate in natural often boil down to robust risk evaluation in one-dimension; see the settings in Sections 6.2, 6.3 and 8.2. If desired by specific applications, the theory of the MA approach can be readily extended to a multi-dimensional setting (see e.g., Embrechts and Puccetti (2006)) with the help from multivariate stochastic orders (e.g., Shaked and Shanthikumar (2007)) and set-valued risk measures (e.g., Hamel and Heyde (2010), Hamel et al. (2011) and Ararat et al. (1999)).

In addition to the multi-dimensional extension mentioned above, we mention a few promising directions of future study. First, one can consider the recently introduced notions of fractional stochastic dominance of Müller et al. (2017) and Huang et al. (2020), which generalize the first- and second-order stochastic dominance used in this paper. Second, instead of relying only on the set $\mathcal{F}$ of uncertainty, which treats each cdf as an element of equal importance ex ante, we can equip a prior probability measure μ on set $\mathcal{F}$, and this will open up many new challenges or conceptualizing and constructing robust models in a similar framework to our theory. Third, we can apply the MA approach to many other settings of uncertainty other than the ones studied in Section 6, and this will lead to convenient tools in various new applications and contexts.

References

Ararat, Ç., Hamel, A. H. and Rudloff, B. (2017). Set-valued shortfall and divergence risk measures. *International Journal of Theoretical and Applied Finance*, **20**(5), 1750026.

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

BCBS (2019). *Minimum Capital Requirements for Market Risk. February 2019.* Basel Committee on Banking Supervision. Basel: Bank for International Settlements.

Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014). Generalized quantiles as risk measures. *Insurance: Mathematics and Economics*, **54**(1), 41–48.

Ben-Tal, A. and Nemirovski, A. (2001). *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. SIAM, Philadelphia, USA.

Blanchet, J., Chen, L. and Zhou, X. (2022). Distributionally robust mean-variance portfolio selection with Wasserstein distances. *Management Science*, **68**(9), 6382–6410.

Blanchet, J., Kang, Y. and Murthy, K. (2019). Robust Wasserstein profile inference and applications to machine learning. *Journal of Applied Probability*, **56**(3), 830–857.

Blanchet, J. and Murthy, K. (2019). Quantifying distributional model risk via optimal transport. *Mathematics of Operations Research*, **44**(2), 565–600.

Cerreia-Vioglio, S., Hansen, L. P., Maccheroni, F. and Marinacci, M. (2021). Making decisions under model misspecification. *Review of Economic Studies*, forthcoming.

Chambers, C. P. (2009). An axiomatization of quantiles on the domain of distribution functions. *Mathematical Finance*, **19**(2), 335–342.

Chen, L., He, S. and Zhang, S. (2011). Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, **59**, 847–865.

Cherny, A. S. and Madan, D. (2009). New measures for performance evaluation. *Review of Financial Studies*, **22**(7), 2571–2606.

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

Cont, R., Deguest, R. and Scandolo, G. (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance*. **10**, 593–606.

Cowell, F. A. (2011). *Measuring Inequality*. 3rd ed. Oxford: Oxford University Press.

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.

Delage, E. and Ye, Y. (2010). Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, **58**(3), 595–612.

Dentcheva, D. (2023). Relations between risk-averse models in extended two-stage stochastic optimization. *Serdica Mathematical Journal*, **49**(1–3), 49–76.

Dentcheva, D. and Ruszczyński, A. (2003). Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, **14**(2), 548–566.

Dentcheva, D. and Ruszczyński, A. (2004). Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints. *Mathematical Programming*, **99**(2), 329–350.

El Ghaoui, L. M. E., Oks, M. and Oustry, F. (2003). Worst-case value-at-risk and robust portfolio optimization. *Operations Research*, **51**, 543–556.

Embrechts, P. and Puccetti, G. (2006). Bounds for functions of multivariate risks. *Journal of Multivariate Analysis*, **97**(2), 526–547.

Embrechts, P., Puccetti, G. and Rüschendorf, L. (2013). Model uncertainty and VaR aggregation. *Journal of Banking and Finance*, **37**(8), 2750–2764.

Embrechts, P., Schied, A. and Wang, R. (2022). Robustness in the optimization of risk measures. *Operations Research*, **70**(1), 95–110.

Embrechts, P., Mao, T., Wang, Q. and Wang, R. (2021). Bayes risk, elicitability, and the Expected Shortfall. *Mathematical Finance*, **31**, 1190–1217.

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), 115–166.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* (Fourth Edition.) Walter de Gruyter, Berlin.

Fu, M., Li, X. and Zhang, L., (2021) Data-driven feature-based newsvendor: A distributionally robust approach. *SSRN*: 3885663.

Gao, R. and Kleywegt, A. J. (2022). Distributionally robust stochastic optimization with Wasserstein distance. *Mathematics of Operations Research*, **48**(2), 603–655.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, **14**(1), 29–58.

Hamel, A. H. and Heyde, F. (2010). Duality for set-valued measures of risk. *SIAM Journal on Financial Mathematics*, **1**(1), 66–95.

Hamel, A. H., Heyde, F. and Rudloff, B. (2011). Set-valued risk measures for conical market models. *Mathematics and Financial Economics*, **5**(1), 1–28.

Han, X., Wang, B., Wang, R. and Wu, Q. (2024). Risk concentration and the mean-Expected Shortfall criterion. *Mathematical Finance*, **34**(3), 819–846.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *American Economic Review*. **91**(2), 60–66.

He, X. and Peng, X. (2018). Surplus-invariant, law-invariant, and conic acceptance sets must be the sets induced by Value-at-Risk. *Operations Research*, **66**(5), 1268–1276.

Huang, R. J., Tzeng, L. Y. and Zhao, L. (2020). Fractional degree stochastic dominance. *Management Science*, **66**(10), 4630–4647.

Jagannathan, R. (1977). Minimax procedure for a class of linear programs under uncertainty. *Operations Research*, **25**(1), 173–177.

Kertz, R. P. and Rösler, U. (2000). Complete lattices of probability measures with applications to martingale theory. *IMS Lecture Notes Monograph Series*, **35**, 153-177.

Klibanoff, P., Marinacci, M. and Mukerji, S. (2005). A smooth model of decision making under uncertainty. *Econometrica*, **73**(6), 1849–1892.

Kou, S. and Peng, X. (2016). On the measurement of economic tail risk. *Operations Research*, **64**(5), 1056–1072.

Kupper, M. and Zapata, J. M. (2021). Large deviations built on max-stability. *Bernoulli*, **27**(2), 1001–1027.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics*. **3**, 83–95.

Leitner, J. (2005). A short note on second-order stochastic dominance preserving coherent risk measures. *Mathematical Finance*, **15**(4), 649–651.

Li, L., Shao, H., Wang, R. and Yang, J. (2018). Worst-case range value-at-risk with partial information. *SIAM Journal on Financial Mathematics*. **9**(1), 190–218.

Li, Y. M. J. (2018). Closed-form solutions for worst-case law invariant risk measures with application to robust portfolio optimization. *Operations Research*, **66**(6), 1533–1541.

Liu, F., Mao, T., Wang, R. and Wei, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519.

Liu, F. and Wang, R. (2021). A theory for measures of tail risk. *Mathematics of Operations Research*, **46**(3), 1109–1128.

Lo, A. (1987). Semiparametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics*, **19**(2), 373–388.

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

Maccheroni, F., Marinacci, M. and Rustichini, A. (2006). Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica*, **74**(6), 1447–1498.

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

Müller, A., Scarsini, M., Tsetlin, I. and Winkler. R. L. (2017). Between first and second-order stochastic dominance. *Management Science*, **63**(9), 2933–2947.

Müller, A. and Scarsini, M. (2006). Stochastic order relations and lattices of probability measures. *SIAM Journal on Optimization*. **16**(4), 1024–1043.

Müller, A. and Stoyan, D. (2002). *Comparison Methods for Stochastic Models and Risks*. John Wiley & Sons, Hoboken.

Natarajan, K., Pachamanova, D. and Sim, M. (2008). Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science*, **54**(3), 573–585.

Natarajan, K., Sim, M. and Uichanco, J. (2010). Tractable robust expected utility and risk models for portfolio optimization. *Mathematical Finance*, **20**, 695–731.

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

Olivares-Nadal, A. V. and DeMiguel, V. (2018). A robust perspective on transaction costs in portfolio optimization. *Operations Research*, **66**(3), 733–739.

Pflug, G. C. (2000). Some remarks on the value-at-risk and the conditional value-at-risk. *Probabilistic Constrained Optimization: Methodology and Applications*, 272–281.

Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. *Operations Research*, **55**(1), 98–112.

Postek, K., den Hertog, D. and Melenberg, B. (2016). Computationally tractable counterparts of distributionally robust constraints on risk measures. *SIAM Review*, **58**(4), 603–650.

Rockafellar, R.T. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance*, **26**(7), 1443–1471.

Rothschild, M. and Stiglitz, J. (1970). Increasing risk I. A definition. *Journal of Economic Theory*, **2**(3), 225–243.

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

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*. Springer Series in Statistics.

Shapiro, A. (2013). On Kusuoka representation of law invariant risk measures. *Mathematics of Operations Research*, **38**(1), 142-152.

Shapiro, A., Dentcheva, D. and Ruszczynski, A. (2021). Lectures on stochastic programming: modeling and theory. *Society for Industrial and Applied Mathematics*.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, **57**(3), 571–587.

Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics*, **17**(1), 43–54.

Wang, R., Peng, L. and Yang, J. (2013). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Finance and Stochastics*, **17**(2), 395–417.

Wang, R., Wei, Y. and Willmot, G. E. (2020). Characterization, robustness and aggregation of signed Choquet integrals. *Mathematics of Operations Research*, **45**(3), 993–1015.

Wang, R. and Ziegel, J. F. (2021). Scenario-based risk evaluation. *Finance and Stochastics*, **25**, 725–756.

Wang, R. and Zitikis, R. (2021). An axiomatic foundation for the Expected Shortfall. *Management Science*, **67**, 1413–1429.

Wiesemann, W., Kuhn, D. and Sim, M. (2014). Distributionally robust convex optimization. *Operations Research*, **62**(6), 1203–1466.

Wozabal, D. (2014). Robustifying convex risk measures for linear portfolios: A nonparametric approach. *Operations Research*, **62**(6), 1302–1315.

Yaari, M. E. (1987). The dual theory of choice under risk. *Econometrica*, **55**(1), 95–115.

Zhu, M. S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155–1168.

Online Supplement: Technical Appendices

Model Aggregation for Risk Evaluation and Robust Optimization

We organize the appendices as follows. We first introduce some extra notation and terminology in Appendix A. Appendix B contains a lemma formally connecting increasing convex order and increasing concave order, and related discussions in Section 3. In Appendix C, we formally introduce the lattice theory and prove a generalized version of Proposition 1 in Section 4. The proofs of the main results are presented in Appendices D (other results in Section 4), E (results in Section 5), F (results and omitted figures in Section 6) and G (results in Section 7). In Appendix H, we present the summary statistics of the return rates as well as the numerical results in portfolio selection under mean-variance uncertainty and Wasserstein uncertainty with a normal benchmark distribution, which complements the numerical studies in Section 8.

A Setting and notation

We will use the same notation as in the main paper. In addition, let L^p be the space of random variables in $(\Omega, \mathcal{B}, \mathbb{P})$ with finite pth moment, $p \in [0, \infty)$, and L^∞ be the space of all bounded random variables. Accordingly, denote by $\mathcal{M}_p$, $p \in [0, \infty]$, the set of cdfs of all random variables in L^p, i.e., $\mathcal{M}_p$ is the set of all cdfs F satisfying $\int_{\mathbb{R}} |x|^p \mathrm{d}F(x) < \infty$ for $p \in [0, \infty)$, and $\mathcal{M}_\infty$ is the set of all compactly supported cdfs. On $\mathcal{M}_\infty$, we can define VaR_0, VaR_1 and ES_1 which are finite, by

$$\mathrm{VaR}_0(F) = \inf\{x \in \mathbb{R} : F(x) > 0\} \ \text{ and } \ \mathrm{VaR}_1(F) = \inf\{x \in \mathbb{R} : F(x) \geqslant 1\}, \quad F \in \mathcal{M}_\infty,$$

and

$$\mathrm{ES}_1(F) = \mathrm{VaR}_1(F), \quad F \in \mathcal{M}_\infty.$$

Denote by $\mathcal{M}_{\mathrm{bb}}$ the set of all cdfs F with a support bounded from below, i.e., $F(x_0) = 0$ for some $x_0 \in \mathbb{R}$. For two real objects (numbers or functions) f and g, $f \vee g$ is their (point-wise) maximum, and $f \wedge g$ is their (point-wise) minimum.

For $p \in [0, \infty]$, the following are some properties of a risk measure $\rho : \mathcal{M}_p \to \mathbb{R}$ and its associated $\widetilde{\rho} : L^p \to \mathbb{R}$. *Translation invariance*: $\widetilde{\rho}(X + c) = \widetilde{\rho}(X) + c$ for any $c \in \mathbb{R}$ and $X \in L^p$. *Positive homogeneity*: $\widetilde{\rho}(\lambda X) = \lambda \widetilde{\rho}(X)$ for any $\lambda > 0$ and $X \in L^p$. *Convexity*: $\widetilde{\rho}(\lambda X + (1 - \lambda)Y) \leqslant \lambda \widetilde{\rho}(X) + (1 - \lambda)\widetilde{\rho}(Y)$ for any $\lambda \in [0, 1]$ and $X, Y \in L^p$. *Lower semicontinuity*: $\liminf_{n\to\infty} \widetilde{\rho}(X_n) \geqslant \widetilde{\rho}(X)$ if $X_n, X \in L^p$ for all n and $X_n \xrightarrow{\mathrm{d}} X$ as $n \to \infty$, where $\xrightarrow{\mathrm{d}}$ denotes convergence in distribution. All the properties are defined for both ρ and $\widetilde{\rho}$.

B Increasing convex order and increasing concave order

For two random variables X and Y in L^1, with cdfs F_X and F_Y, respectively, we define the *second-order stochastic dominance* $\preceq_{\rm ssd}$, also known as the *increasing concave order*, as $F_X \preceq_{\rm ssd} F_Y$ if $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all increasing concave functions u. Here, we define $\preceq_{\rm ssd}$ and $\preceq_2$ for both random variables and their cdfs; that is, we write $X \preceq_2 Y$ for $F_X \preceq_2 F_Y$, and similarly for $\preceq_{\rm ssd}$. The following lemma clarifies the connection between these two partial orders.

Lemma EC.1. *For any random variables X and Y in L^1, we have $X \preceq_{\rm ssd} Y$ if and only if $-Y \preceq_2 -X$. Moreover, for a given $Y \in L^1$, the set $\{X \in L^1 : X \preceq_2 Y\}$ is convex.*

Proof. Note that u is increasing convex if and only if $x \mapsto -u(-x)$ is increasing concave. The equivalence condition between the two partial orders follows from this observation. To be more specific, let $U_{\rm icx}$ (resp. $U_{\rm icv}$) be the set of all increasing convex (resp. concave) functions on $\mathbb{R}$. It holds that $U_{\rm icv} = \{v : v(x) = -u(-x),\, x \in \mathbb{R},\, u \in U_{\rm icx}\}$. Therefore, we have $X \preceq_{\rm ssd} Y$, that is, $\mathbb{E}[u(X)] \leqslant \mathbb{E}[u(Y)]$ for all $u \in U_{\rm icv}$ if and only if $\mathbb{E}[u(-X)] \geqslant \mathbb{E}[u(-Y)]$ for all $u \in U_{\rm icx}$, that is, $-Y \preceq_2 -X$.

Based on this equivalence condition, we have for a given $Y \in L^1$, it holds that $\{X \in L^1 : X \preceq_2 Y\} = \{X \in L^1 : -Y \preceq_{\rm ssd} -X\}$. The convexity of this set follows from Dentcheva and Ruszczynski (2003, Proposition 2.3), which states that for any $Z \in L^1$, $\{X \in L^1 : Z \preceq_{\rm ssd} X\}$ is convex, and thus, $\{X \in L^1 : -Y \preceq_{\rm ssd} X\} = \{-X \in L^1 : -Y \preceq_{\rm ssd} -X\}$ is convex. This completes the proof. □

It is worth noting that Dentcheva and Ruszczynski (2004) used $\preceq_{\rm ssd}$ in the constraints of their optimization problems, whereas we use $\preceq_2$ in Problems (8) and (9). In the following, we explain why our formulation remains consistent with that of Dentcheva and Ruszczynski (2004).

Let $\mathbf{Z}$ be a given random vector and each of its positive components represents a gain; Given functions u, $h_1, \dots, h_m$ and $G_1, \dots, G_m \in \mathcal{M}_1$, Dentcheva and Ruszczynski (2004) considered the following optimization problem

$$\max_{\mathbf{a}\in A} \ \ \mathbb{E}[u(\mathbf{a}, \mathbf{Z})] \qquad \text{s.t. } F_{h_i(\mathbf{a},\mathbf{Z})} \succeq_{\rm ssd} G_i, \ \ i \in [m], \tag{EC.1}$$

where A is a set of possible actions. Note that we consider random losses, denoted by $\mathbf{X}$, which represent the negative value of the gains, i.e., $\mathbf{X} = -\mathbf{Z}$. Define $f(\mathbf{a}, \mathbf{x}) = -u(\mathbf{a}, -\mathbf{x})$ and $g_i(\mathbf{a}, \mathbf{x}) = -h_i(\mathbf{a}, -\mathbf{x})$, $i \in [m]$. Therefore, we have

$$f(\mathbf{a}, \mathbf{X}) = -u(\mathbf{a}, -\mathbf{X}) = -u(\mathbf{a}, \mathbf{Z}) \ \text{ and } \ g_i(\mathbf{a}, \mathbf{X}) = -h_i(\mathbf{a}, -\mathbf{X}) = -h_i(\mathbf{a}, \mathbf{Z}), \quad i \in [m].$$

Further, let F_i be the distribution of $-Y_i$, where Y_i has distribution G_i for $i \in [m]$. Then, it follows from Lemma EC.1 that the constraint in (EC.1) is equivalent to $F_{g_i(\mathbf{a},\mathbf{X})} \preceq_2 F_i$, $i \in [m]$. Therefore, Problem (EC.1) is equivalent to

$$\min_{\mathbf{a}\in A} \ \mathbb{E}[f(\mathbf{a},\mathbf{X})] \qquad \text{s.t. } F_{g_i(\mathbf{a},\mathbf{X})} \preceq_2 F_i, \quad i \in [m],$$

which is exactly our Problem (8).

C Lattice theory and the proof of Proposition 1

In this appendix, we introduce the lattice structure of an ordered set which complements the main paper. For more details of the lattice theory, the reader is referred to Davey and Priestley (2002).

Definition EC.1. Let $(\mathcal{M}, \preceq)$ be an ordered set (i.e., $\preceq$ is a partial order on $\mathcal{M}$) and $\mathcal{F} \subseteq \mathcal{M}$.

(i) A set $\mathcal{F}$ is said to be *bounded from above* (*below*, resp.) in $(\mathcal{M}, \preceq)$, if the set of upper (lower, resp.) bounds on $\mathcal{F}$, denoted by $U(\mathcal{F})$ ($L(\mathcal{F})$, resp.), is nonempty, where

$$U(\mathcal{F}) = \{G \in \mathcal{M} : F \preceq G, \ \forall F \in \mathcal{F}\} \ \text{ and } \ L(\mathcal{F}) = \{G \in \mathcal{M} : G \preceq F, \ \forall F \in \mathcal{F}\}. \quad \text{(EC.2)}$$

(ii) For $\mathcal{F} \subseteq \mathcal{M}$ which is bounded from above (below, resp.), if there exists $F_0 \in U(\mathcal{F})$ ($L(\mathcal{F})$, resp.) such that $F_0 \preceq$ ($\succeq$, resp.) G for all $G \in U(\mathcal{F})$ ($L(\mathcal{F})$, resp.), then F_0 is called the *supremum* (*infimum*, resp.) of $\mathcal{F}$ and we write $\bigvee \mathcal{F} = F_0$ ($\bigwedge \mathcal{F} = F_0$, resp.).

(iii) If for all $F, G \in \mathcal{M}$, $\bigvee\{F, G\}$ and $\bigwedge\{F, G\}$ exist, then $(\mathcal{M}, \preceq)$ is called a *lattice*. If $\bigvee \mathcal{F}$ exists for all $\mathcal{F} \subseteq \mathcal{M}$ that are bounded from above and $\bigwedge \mathcal{F}$ exists for all $\mathcal{F} \subseteq \mathcal{M}$ that are bounded from below, then $(\mathcal{M}, \preceq)$ is called a *complete lattice*.[15]

Remark EC.1. In case $(\mathcal{M}, \preceq)$ is a lattice which is not complete, $\bigvee \mathcal{F}$ may not exist even if $\mathcal{F}$ is bounded from above. In this case, the definition of the MA robust risk value needs to be modified. We can alternatively define $\rho^{\mathrm{MA}}(\mathcal{F}) = \inf_{G \in U(\mathcal{F})} \rho(G)$ where $U(\mathcal{F})$ is defined by (EC.2), and this definition is equivalent to (2) if $(\mathcal{M}, \preceq)$ is a complete lattice and ρ is $\preceq$-consistent. For any partial order $\preceq$, the supremum and infimum are both unique whenever they exist.

[15] The definition of complete lattice in Davey and Priestley (2002) is slightly different to ours. In Davey and Priestley (2002), a complete lattice has the largest and the smallest elements, and our $\mathcal{M}$ does not. Nevertheless, if we extend $\mathcal{M}$ to $\overline{\mathcal{M}} := \mathcal{M} \cup \{F_{\min}, F_{\max}\}$ where $F_{\min} \preceq F$ and $F \preceq F_{\max}$ for all $F \in \mathcal{M}$, then our definition of complete lattice on the ordered set $(\overline{\mathcal{M}}, \preceq)$ is equivalent to the one of Davey and Priestley (2002).

For stochastic dominances $\preceq_1$ and $\preceq_2$, there are several equivalent definitions that are useful throughout the paper; see e.g., Bäuerle and Müller (2006). In case of $\preceq_1$, the following statements are equivalent: (i) $F \preceq_1 G$; (ii) $F(x) \geqslant G(x)$ for all $x \in \mathbb{R}$; (iii) $F^{-1}(\alpha) \leqslant G^{-1}(\alpha)$ for all $\alpha \in (0,1)$. In case of $\preceq_2$, the following statements are equivalent: (i) $F \preceq_2 G$; (ii) $\pi_F(x) \leqslant \pi_G(x)$ for all $x \in \mathbb{R}$ where π_F is the integrated survival function defined by (10); (iii) $E_F(\alpha) \leqslant E_G(\alpha)$ for all $\alpha \in (0,1)$ where E_F is the *integrated quantile function*[16] defined by

$$E_F(\alpha) = (1-\alpha)\mathrm{ES}_\alpha(F) = \int_\alpha^1 F^{-1}(s)\mathrm{d}s, \quad \alpha \in [0,1]. \tag{EC.3}$$

The complete lattice structure of $(\mathcal{M}_0, \preceq_1)$ and $(\mathcal{M}_+, \preceq_2)$ and the formulas for the suprema are known in the literature; see Kertz and Rösler (2000). Here $\mathcal{M}_+ = \{F \in \mathcal{M}_0 : \int_0^\infty x\mathrm{d}F(x) < \infty\}$. The following proposition which is a generalized result of Proposition 1 considers general space $\mathcal{M}_p$, $p \in [0,\infty]$ with partial order $\preceq_1$ and $\preceq_2$. Specifically, in Proposition EC.1 below, (a) generalizes Proposition 1 (a) to the domain $\mathcal{M}_p$, $p \in [0,\infty]$, and similarly, (b) and (c) generalize Proposition 1 (b).

Proposition EC.1. *(a) For each $p \in [0,\infty]$, the partially ordered set $(\mathcal{M}_p, \preceq_1)$ is a complete lattice. If $\mathcal{F} \subseteq \mathcal{M}_p$ is bounded from above, then its supremum $\bigvee_1 \mathcal{F}$ is given by $\inf_{F\in\mathcal{F}} F$, and the left quantile function of $\bigvee_1 \mathcal{F}$ is $\sup_{F\in\mathcal{F}} F^{-1}$.*

(b) The partially ordered set $(\mathcal{M}_1, \preceq_2)$ is a complete lattice, and for $\mathcal{F}$ that is bounded from above,

$$\pi_{\bigvee_2 \mathcal{F}} = \sup_{F\in\mathcal{F}} \pi_F, \qquad \bigvee_2 \mathcal{F} = 1 + \left(\sup_{F\in\mathcal{F}} \pi_F\right)'_+.$$

(c) For each $p \in (1,\infty]$, the ordered set $(\mathcal{M}_p, \preceq_2)$ is a lattice and not a complete lattice. The supremum is given by $\bigvee_2\{F,G\} = 1 + (\pi_F \vee \pi_G)'_+$ for $F, G \in \mathcal{M}_p$.

Proof. We first give one fact: For $p \in [0,\infty]$ and an increasing and right-continuous function $H : \mathbb{R} \to [0,1]$, if $F, G \in \mathcal{M}_p$ and $F \leqslant H \leqslant G$, then $H \in \mathcal{M}_p$. It suffices to verify that

1. $0 \leqslant \lim_{x\to-\infty} H(x) \leqslant \lim_{x\to-\infty} G(x) = 0$ and $1 \geqslant \lim_{x\to\infty} H(x) \geqslant \lim_{x\to-\infty} F(x) = 1$, which imply that H is a cdf on $\mathbb{R}$, that is, $H \in \mathcal{M}_0$.

2. If $p \in (0,\infty)$, then we have $F \succeq_1 H \succeq_1 G$ and thus $\int_0^\infty x^p\mathrm{d}H(x) \leqslant \int_0^\infty x^p\mathrm{d}F(x) < \infty$ and $\int_{-\infty}^0 (-x)^p\mathrm{d}H(x) \leqslant \int_{-\infty}^0 (-x)^p\mathrm{d}G(x) < \infty$. It follows that $\int_{\mathbb{R}} |x|^p\mathrm{d}H(x) < \infty$, that is, $H \in \mathcal{M}_p$.

[16] The integrated quantile function E_F is also called the (upper) absolute Lorenz function, see, e.g., Shorrocks (1983) and Cowell (2011).

3. If $F, G \in \mathcal{M}_\infty$, then there exists x, $y \in \mathbb{R}$ such that $G(x) = 0$ and $F(y) = 1$. Then we have $H(x) = 0$ and $H(y) = 1$, that is, $H \in \mathcal{M}_\infty$.

(a) For $p \in [0, \infty]$, let $\mathcal{F} \subseteq \mathcal{M}_p$. Suppose that $\mathcal{F}$ is bounded from above. Define $H = \inf_{F\in\mathcal{F}} F$ which is increasing and right-continuous. Then there exists $G \in \mathcal{M}_p$ such that $F \geqslant H \geqslant G$ for any $F \in \mathcal{F}$. By the above fact, we have $H \in \mathcal{M}_p$. If $\mathcal{F}$ is bounded from below, define $H(x) = \lim_{y\downarrow x} H_1(y)$ where $H_1 = \sup_{F\in\mathcal{F}} F$. Then H is increasing and right-continuous and there exists $G \in \mathcal{M}_p$ such that $G \geqslant H \geqslant F$ for any $F \in \mathcal{F}$. By the above fact, we have $H \in \mathcal{M}_p$. Therefore, we have that $(\mathcal{M}_p, \preceq_1)$ is a complete lattice for $p \in [0, \infty]$. The statement on the left quantile of $\bigvee_1 \mathcal{F}$ follows from $(\inf_{F\in\mathcal{F}} F)^{-1} = \sup_{F\in\mathcal{F}} F^{-1}$. Hence, we complete the proof of (a).

(b) The proof is similar to that of Theorem 3.4 of Kertz and Rösler (2000) which shows that $(\mathcal{M}_+, \preceq_2)$ is a complete lattice. We give a proof for completeness. Let $\mathcal{F} \subseteq \mathcal{M}_1$ be bounded from above. There exists $G \in \mathcal{M}_1$ such that $F \preceq_2 G$ for all $F \in \mathcal{F}$, that is, $\sup_{F\in\mathcal{F}} \pi_F(x) \leqslant \pi_G(x)$ for $x \in \mathbb{R}$. One can check that

1. $\pi_0(x) := \sup_{F\in\mathcal{F}} \pi_F(x)$ is decreasing convex as each $\pi_F(x)$ is decreasing convex. This implies $1 + (\pi_0)'_+(x)$ is right-continuous and increasing.

2. $\lim_{x\to\infty} \pi_0(x) \leqslant \lim_{x\to\infty} \pi_G(x) = 0$ which implies $\lim_{x\to\infty} (\pi_0)'_+(x) = 0$, that is, $\lim_{x\to\infty}(1 + (\pi_0)'_+(x)) = 1$.

3. Since $x + \pi_F(x)$ is increasing in x for all $F \in \mathcal{F}$, we have $x + \pi_0(x)$ is increasing in x and thus $\lim_{x\to-\infty} x + \pi_0(x)$ exists (may take $-\infty$). Let $F^* \in \mathcal{F}$, and we have $x + \pi_0(x) \geqslant x + \pi_{F^*}(x)$ for all $x \in \mathbb{R}$. Noting that $\lim_{x\to-\infty} x + \pi_{F^*}(x) = \mathfrak{m}(F^*) \in \mathbb{R}$, we have $\lim_{x\to-\infty} x + \pi_0(x) \in \mathbb{R}$, which implies $\lim_{x\to-\infty} 1 + (\pi_0)'_+(x) = 0$.

Combining the above three observations, we have $H = 1 + (\sup_{F\in\mathcal{F}} \pi_F)'_+ \in \mathcal{M}_1$. By definition of supremum, it is standard to check that $\bigvee_2 \mathcal{F} = H$.

Let $\mathcal{F} \subseteq \mathcal{M}_1$ be bounded from below. There exists $G \in \mathcal{M}_1$ such that $G \preceq_2 F$ for all $F \in \mathcal{F}$, that is, $E_G(\alpha) \leqslant \inf_{F\in\mathcal{F}} E_F(\alpha)$ for $\alpha \in [0, 1]$. Similar to the proof of Steps 1-3 for $\mathcal{F}$ that is bounded from above, one can show that $\inf_{F\in\mathcal{F}} E_F$ is an integrated quantile function of some cdf in $\mathcal{M}_1$, say H. By definition of infimum, we have $H = \bigwedge_2 \mathcal{F}$. It follows from the relation between a cdf and its integrated quantile function that $H^{-1} = -(\inf_{F\in\mathcal{F}} E_F)'_-$. This completes the proof of (b).

(c) For $F, G \in \mathcal{M}_p$, define $F_1 = \bigvee_2\{F, G\}$ and $F_2 = \bigwedge_2\{F, G\}$. It follows from (b) that $F_1 = 1 + (\pi_F \vee \pi_G)'_+$ which implies $\min\{F, G\} \leqslant F_1 \leqslant \max\{F, G\}$, and $F_2^{-1} = -(E_F \wedge E_G)'_-$ which implies $\min\{F^{-1}, G^{-1}\} \leqslant F_2^{-1} \leqslant \max\{F^{-1}, G^{-1}\}$, and hence, $\min\{F, G\} \leqslant F_2 \leqslant \max\{F, G\}$. By the fact in the beginning of the proof, we have $F_1, F_2 \in \mathcal{M}_p$, and thus $(\mathcal{M}_p, \preceq_2)$ is a lattice for $p \in (1, \infty]$.

Below, we give a counterexample to illustrate that $(\mathcal{M}_p, \preceq_2)$ is not complete lattice for $p \in (1, \infty]$. For $p \in (1, \infty)$, define $F(x) = (-x)^{-p}$ for $x \leqslant -1$. We have $F \notin \mathcal{M}_p$ and for $y < -1$, let F_y be a cdf with integrated survival function

$$\pi_{F_y}(x) = \max\left\{\left(-x - \frac{p}{p-1}\right)_+,\ \pi'_F(y)(x-y) + \pi_F(y)\right\}.$$

It is clear that $F_y \in \mathcal{M}_\infty$ for all $y < -1$ and the set $\{F_y\}_{y<-1}$ is bounded from above as $F_y \preceq_2 \delta_{-1}$ for $y < -1$. Noting that $\sup_{y<-1} \pi_{F_y} = \pi_F$ and $F \notin \mathcal{M}_p$, we have that $(\mathcal{M}_p \preceq_2)$ is not a complete lattice. $\square$

D Proofs for other results in Sections 4

Proof of Proposition 2. For a fixed $x \in \mathbb{R}$, both $F \mapsto F(x)$ and $F \mapsto \pi_F(x)$ are affine on $\mathcal{M}_1$. Hence, for $F \in \mathrm{conv}\mathcal{F}$ with $F = \sum_{i=1}^n \lambda_i F_i$ where $(\lambda_1, \dots, \lambda_n) \in \Delta_n$ and $F_i \in \mathcal{F}$ for $i \in [n]$, there exist $G_1, G_2 \in \{F_1, \dots, F_n\} \subseteq \mathcal{F}$ such that $G_1(x) \leqslant F(x)$ and $\pi_{G_2}(x) \geqslant \pi_F(x)$. The results follow immediately from Proposition 1. $\square$

Proof of Theorem 1. (a) Since $\mathfrak{m}(F) = \lim_{x\to-\infty}\{x + \pi_F(x)\}$ for each $F \in \mathcal{M}_1$, we have

$$\begin{aligned}
\mathfrak{m}^{\mathrm{MA}_2}(\mathcal{F}) &= \mathfrak{m}\left(\bigvee_2 \mathcal{F}\right) = \lim_{x\to-\infty}\left\{x + \pi_{\bigvee_2 \mathcal{F}}(x)\right\} = \lim_{x\to-\infty}\left\{x + \sup_{F\in\mathcal{F}} \pi_F(x)\right\} \\
&= \lim_{x\to-\infty} \sup_{F\in\mathcal{F}} \left\{\mathfrak{m}(F) + \int_{\mathbb{R}} (x-y)_+ \mathrm{d}F(y)\right\} \leqslant \sup_{F\in\mathcal{F}} \mathfrak{m}(F) + \lim_{x\to-\infty} \sup_{F\in\mathcal{F}} \int_{\mathbb{R}} (x-y)_+ \mathrm{d}F(y) \\
&= \sup_{F\in\mathcal{F}} \mathfrak{m}(F) = \mathfrak{m}^{\mathrm{WR}}(\mathcal{F}),
\end{aligned}$$

where the third equality comes from (ii) of Proposition 1, and the forth equality follows from $x + \pi_F(x) = x + \mathbb{E}^F[(X-x)_+] = \mathfrak{m}(F) + \mathbb{E}^F[(x-X)_+]$. The converse direction $\mathfrak{m}^{\mathrm{MA}_2}(\mathcal{F}) \geqslant \mathfrak{m}^{\mathrm{WR}}(\mathcal{F})$ is trivial. Hence, we complete the proof of (a).

(b) Suppose that $\mathcal{F} \subseteq \mathcal{M}_1$ is a convex set which is $\preceq_2$-bounded. Denote by $\Pi_{\mathcal{G}} = \sup_{F\in\mathcal{G}} \pi_F$ for any set $\mathcal{G} \subseteq \mathcal{M}_1$. If $\mathcal{G}$ is a convex polytope, then by Theorem 1 of Zhu and Fukushima (2009), we have

$$\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{G}) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{G}). \tag{EC.4}$$

Let $c = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F})$. Using (14), we get

$$x + \frac{1}{1-\alpha}\Pi_{\mathcal{F}}(x) \geqslant c \ \text{ for all } x \in \mathbb{R}. \tag{EC.5}$$

Take an arbitrary $G \in \mathcal{F}$. Since $(\pi_G)'_+(x) \to -1$ as $x \to -\infty$, we have $(1-\alpha)x + \pi_G(x) \to \infty$ as $x \to -\infty$. There exists $x_0 < c$ such that

$$x + \frac{1}{1-\alpha}\pi_G(x) \geqslant c \ \text{ for all } x < x_0. \tag{EC.6}$$

Fix $\varepsilon > 0$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a convex polytope such that

$$\Pi_{\mathcal{G}}(x) \geqslant \Pi_{\mathcal{F}}(x) - \varepsilon \ \text{ for all } x \in [x_0, c]. \tag{EC.7}$$

We illustrate why such $\mathcal{G}$ exists. Let $\mathbb{Q}$ be a set of all rational numbers on $\mathbb{R}$, and we represent it as $\mathbb{Q} = \{q_i\}_{i\in\mathbb{N}}$. Suppose that $\{F_{i,j}\}_{j\in\mathbb{N}} \subseteq \mathcal{F}$ satisfies $\lim_{j\to\infty} \pi_{F_{i,j}}(q_i) = \Pi_{\mathcal{F}}(q_i)$ for $i \in \mathbb{N}$. Define $\mathcal{G}_i = \{F_{1,i}, F_{2,i}, \dots, F_{i,i}\}$ for $i \in \mathbb{N}$. It holds that $\Pi_{\mathcal{G}_i}(x) \to \Pi_{\mathcal{F}}(x)$ on $\mathbb{Q}$. By Theorem 10.8 of Rockafellar (1970), we have $\{\Pi_{\mathcal{G}_i}\}_{i\in\mathbb{N}}$ uniformly converges to $\Pi_{\mathcal{F}}$ on $[x_0, c]$. This implies that such $\mathcal{G}$ in (EC.7) exists. Let $\mathcal{G}_0 = \mathrm{conv}(\mathcal{G} \cup \{G\}) \subseteq \mathcal{F}$, which is again a convex polytope. Using (EC.4), (EC.5), (EC.6) and (EC.7), we obtain

$$\begin{aligned}
\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{G}_0) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{G}_0) &= \min_{x\in\mathbb{R}} \left\{ x + \frac{\Pi_{\mathcal{G}_0}(x)}{1-\alpha} \right\} \\
&= \min\left\{ \inf_{x<x_0} \left\{ x + \frac{\Pi_{\mathcal{G}_0}(x)}{1-\alpha} \right\}, \min_{x\in[x_0,c]} \left\{ x + \frac{\Pi_{\mathcal{G}_0}(x)}{1-\alpha} \right\}, \inf_{x>c} \left\{ x + \frac{\Pi_{\mathcal{G}_0}(x)}{1-\alpha} \right\} \right\} \\
&\geqslant \min\left\{ \inf_{x<x_0} \left\{ x + \frac{\pi_G(x)}{1-\alpha} \right\}, \min_{x\in[x_0,c]} \left\{ x + \frac{\Pi_{\mathcal{G}}(x)}{1-\alpha} \right\}, c \right\} \\
&\geqslant \min\left\{ \min_{x\in[x_0,c]} \left\{ x + \frac{\Pi_{\mathcal{F}}(x)}{1-\alpha} - \frac{\varepsilon}{1-\alpha} \right\}, c \right\} \\
&\geqslant \min\left\{ \min_{x\in\mathbb{R}} \left\{ x + \frac{\Pi_{\mathcal{F}}(x)}{1-\alpha} \right\}, c \right\} - \frac{\varepsilon}{1-\alpha} = c - \frac{\varepsilon}{1-\alpha}.
\end{aligned}$$

Note that $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) \geqslant \mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{G}_0) \geqslant c - \varepsilon/(1-\alpha)$ because $\mathcal{G}_0 \subseteq \mathcal{F}$. Since ε is arbitrary, we get $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) \geqslant c = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F})$. Together with $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) \leqslant \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F})$, we obtain the desired equality $\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}) = \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F})$.

(c) It follows directly from Proposition 1. □

Proof of Theorem 2. (a) We first consider the convexity of $\widetilde{\rho}^{\mathrm{WR}}$. Suppose that $\widetilde{\rho}$ is convex. Note that $\widetilde{\rho}^{\mathrm{WR}}$ is the supremum of a family of functionals $\widetilde{\rho}^{Q}$, $Q \in \mathcal{Q}$. It suffices to verify that $\widetilde{\rho}^{Q}$ is convex on L for all $Q \in \mathcal{Q}$. Let $Q \in \mathcal{Q}$ and $X_1, X_2 \in L$. By a version of Skorhod's Theorem (see, e.g., Theorem 3.1 of Berti et al. (2007)), we can construct measurable mappings X_1', X_2' on $(\Omega, \mathcal{B})$ such that $\mathbb{P}(X_1' \leqslant x_1, X_2' \leqslant x_2) = Q(X_1 \leqslant x_1, X_2 \leqslant x_2)$ for all $x_1, x_2 \in \mathbb{R}$, i.e., the joint cdf of (X_1, X_2) under Q is same as the joint cdf of (X_1', X_2') under $\mathbb{P}$. Hence, we have $X_1', X_2' \in L^1$, and

$$\widetilde{\rho}^Q(\lambda X_1 + (1-\lambda)X_2) = \widetilde{\rho}(\lambda X_1' + (1-\lambda)X_2') \leqslant \lambda\widetilde{\rho}(X_1') + (1-\lambda)\widetilde{\rho}(X_2') = \lambda\widetilde{\rho}^Q(X_1) + (1-\lambda)\widetilde{\rho}^Q(X_2),$$

where the inequality follows from the convexity of ρ. This yields the convexity of $\widetilde{\rho}^{\mathrm{WR}}$.

To see the convexity of $\widetilde{\rho}^{\mathrm{MA}_2}$, by Theorem 2.2 of Kaina and Rüschendorf (2009), $\widetilde{\rho}: L^1 \to \mathbb{R}$ is continuous with respect to the L^1-norm because $\widetilde{\rho}(X) \in \mathbb{R}$ for all $X \in L^1$. It then follows from Theorem 5.1 of Cerreia-Vioglio et al. (2011) that ρ is $\preceq_2$-consistent. For any $F \in \mathcal{M}_1$, denote $g_F(\alpha) = \mathrm{ES}_\alpha(F)$ for $\alpha \in (0,1)$, and define

$$\mathcal{G} = \{g_F : (0,1) \to \mathbb{R} \mid F \in \mathcal{M}_1\} \quad \text{and} \quad \widehat{\rho}: \mathcal{G} \to \mathbb{R} \;\text{ as }\; \widehat{\rho}(g_F) = \rho(F), \quad g_F \in \mathcal{G}.$$

We assert that $\mathcal{G}$ is a convex set and $\widehat{\rho}$ is convex in $\mathcal{G}$. To see it, take $F, G \in \mathcal{M}_1$ and $\lambda \in [0,1]$. Define H as a distribution whose quantile is $H^{-1}(\alpha) = \lambda F^{-1}(\alpha) + (1-\lambda)G^{-1}(\alpha)$, $\alpha \in [0,1]$. One can verify that $H \in \mathcal{M}_1$ and $g_H = \lambda g_F + (1-\lambda)g_G$, and thus, $\lambda g_F + (1-\lambda)g_G \in \mathcal{G}$, which implies that $\mathcal{G}$ is a convex set. Let $U \in L^1$ be a uniform random variable on $[0,1]$ under $\mathbb{P}$, i.e., $\mathbb{P}(U \leqslant x) = x$ for $x \in [0,1]$, and such random variable exists because $(\Omega, \mathcal{F}, \mathbb{P})$ is a nonatomic space (see e.g., Lemma A.27 of Föllmer and Schied (2016)). It holds that $\lambda F^{-1}(U) + (1-\lambda)G^{-1}(U) \in L^1$ has the distribution H. Therefore, we have

$$\begin{aligned}\widehat{\rho}(\lambda g_F + (1-\lambda)g_G) = \widehat{\rho}(g_H) &= \rho(H)\\ &= \widetilde{\rho}(\lambda F^{-1}(U) + (1-\lambda)G^{-1}(U)) \leqslant \lambda\widetilde{\rho}(F^{-1}(U)) + (1-\lambda)\rho(G^{-1}(U))\\ &= \lambda\rho(F) + (1-\lambda)\rho(G) = \lambda\widehat{\rho}(g_F) + (1-\lambda)\widehat{\rho}(g_G),\end{aligned}$$

where we have used the convexity of $\widetilde{\rho}$ in the first inequality. This implies that $\widehat{\rho}$ is convex in $\mathcal{G}$. We assert that

$$\widetilde{\rho}^{\mathrm{MA}_2}(X) = \inf_{g\in\mathcal{G}} \{\widehat{\rho}(g) + \Theta(X,g)\}, \tag{EC.8}$$

where $\Theta(X,g) = 0$ if $\sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(X) \leqslant g(\alpha)$ for all $\alpha \in (0,1)$, and $\Theta(X,g) = \infty$ otherwise. To see it, note that if $\mathcal{F}_{X|\mathcal{Q}}$ is not bounded, then $\widetilde{\rho}^{\mathrm{MA}_2}(X) = \infty$ and $\Theta(X,g) = \infty$ for all $g \in \mathcal{G}$, which imply (EC.8) holds. If $\mathcal{F}_{X|\mathcal{Q}}$ is bounded, then note that $F \preceq_2 G$ if and only if $g_F(\alpha) \leqslant g_G(\alpha)$ for all $\alpha \in (0,1)$. Since ρ is $\preceq_2$-consistent, we have

$$\widetilde{\rho}^{\mathrm{MA}_2}(X) = \rho\left(\bigvee_2 \mathcal{F}_{X|\mathcal{Q}}\right) = \inf\left\{\rho(F) : F \in \mathcal{M}_1,\ \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(X) \leqslant g_F(\alpha),\quad \forall \alpha \in (0,1)\right\}$$
$$= \inf\left\{\widehat{\rho}(g) : g \in \mathcal{G},\ \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(X) \leqslant g(\alpha),\quad \forall \alpha \in (0,1)\right\}$$
$$= \inf_{g\in\mathcal{G}} \left\{\widehat{\rho}(g) + \Theta(X,g)\right\}.$$

Therefore, (EC.8) holds for all $X \in L$. It remains to show that $\widehat{\rho}(g) + \Theta(X,g)$ is convex in $(X,g) \in L \times \mathcal{G}$. Since $\widehat{\rho}$ is convex in $\mathcal{G}$, it remains to verify that $\Theta(X,g)$ is convex on $L \times \mathcal{G}$. To see this, by definition of $\Theta(X,g)$, it suffices to show that the set $\{(X,g) \in L \times \mathcal{G} : \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(X) \leqslant g(\alpha),\ \alpha \in [0,1]\}$ is a convex set. Take $X_1, X_2 \in L$ and $g_1, g_2 \in \mathcal{G}$. For $\alpha \in [0,1]$, it holds that

$$\sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(\lambda X_1 + (1-\lambda)X_2) \leqslant \sup_{Q\in\mathcal{Q}} \left\{\lambda \widetilde{\mathrm{ES}}_\alpha^Q(X_1) + (1-\lambda)\widetilde{\mathrm{ES}}_\alpha^Q(X_2)\right\}$$
$$\leqslant \lambda \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(X_1) + (1-\lambda) \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{ES}}_\alpha^Q(X_2) \leqslant \lambda g_1 + (1-\lambda) g_2,$$

where the first inequality follows from the convexity of ES. Hence, we have that $\widetilde{\rho}^{\mathrm{MA}_2}$ is convex.

(b) Define $f_F(\alpha) = \mathrm{VaR}_\alpha(F)$ for $\alpha \in (0,1)$ and $F \in \mathcal{M}_1$. Let $\mathcal{M}_1^{-1} = \{f_F : (0,1) \to \mathbb{R} | F \in \mathcal{M}_1\}$. Define $\widehat{\rho} : \mathcal{M}_1^{-1} \to \mathbb{R}$ as the risk measure satisfying $\widehat{\rho}(f_F) = \rho(F)$. Noting that $\widetilde{\rho}$ satisfies comonotonic additivity, we have $\widehat{\rho}(f_1 + f_2) = \widehat{\rho}(f_1) + \widehat{\rho}(f_2)$ for all $f_1, f_2 \in \mathcal{M}_1^{-1}$. Suppose now $X_1, X_2 \in L$ are two comonotonic random variables. By Proposition 4.6 of Wang and Ziegel (2021) who prove that the mapping $X \mapsto \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{VaR}}_\alpha^Q(X)$ satisfies comonotonic additivity for all $\alpha \in (0,1)$, we have

$$\sup_{Q\in\mathcal{Q}} f_{F^Q_{X_1+X_2}}(\alpha) = \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{VaR}}_\alpha^Q(X_1 + X_2)$$
$$= \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{VaR}}_\alpha^Q(X_1) + \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{VaR}}_\alpha^Q(X_2) = \sup_{Q\in\mathcal{Q}} f_{F^Q_{X_1}}(\alpha) + \sup_{Q\in\mathcal{Q}} f_{F^Q_{X_2}}(\alpha),\quad \forall \alpha \in (0,1).$$

Hence,

$$\widetilde{\rho}^{\mathrm{MA}_1}(X_1 + X_2) = \rho\left(\bigvee_1 \mathcal{F}_{X_1+X_2|\mathcal{Q}}\right) = \widehat{\rho}\left(\sup_{Q\in\mathcal{Q}} f_{F^Q_{X_1+X_2}}\right)$$
$$= \widehat{\rho}\left(\sup_{Q\in\mathcal{Q}} f_{F^Q_{X_1}} + \sup_{Q\in\mathcal{Q}} f_{F^Q_{X_2}}\right) = \widehat{\rho}\left(\sup_{Q\in\mathcal{Q}} f_{F^Q_{X_1}}\right) + \widehat{\rho}\left(\sup_{Q\in\mathcal{Q}} f_{F^Q_{X_2}}\right)$$
$$= \widetilde{\rho}^{\mathrm{MA}_1}(X_1) + \widetilde{\rho}^{\mathrm{MA}_1}(X_2).$$

This yields the comonotonic additivity of $\widetilde{\rho}^{\mathrm{MA}_1}$.

(c) We only consider the case of translation invariance as the case of positive homogeneity is similar. Suppose that $\widetilde{\rho}$ satisfies translation invariance. We have $\rho(G) = \rho(F) + c$ whenever $F, G \in \mathcal{M}_1$ satisfy $G(x) = F(x - c)$ for all $x \in \mathbb{R}$. The translation invariance of $\widetilde{\rho}^{\mathrm{WR}}$ is trivial because we have

$$\widetilde{\rho}^{Q}(X + c) = \rho(F^{Q}_{X+c}) = \rho(F^{Q}_{X}) + c = \widetilde{\rho}^{Q}(X) + c$$

for any $X \in L$, $Q \in \mathcal{Q}$ and $c \in \mathbb{R}$. To see the case of MA_1, it follows from Proposition 1 that

$$\mathrm{VaR}_\alpha\left(\bigvee_1 \mathcal{F}_{X+c|\mathcal{Q}}\right) = \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{VaR}}^{Q}_\alpha(X+c) = \sup_{Q\in\mathcal{Q}} \widetilde{\mathrm{VaR}}^{Q}_\alpha(X)+c = \mathrm{VaR}_\alpha\left(\bigvee_1 \mathcal{F}_{X|\mathcal{Q}}\right)+c, \ \ \forall \alpha \in (0,1).$$

This means that $\bigvee_1 \mathcal{F}_{X+c|\mathcal{Q}}(x) = \bigvee_1 \mathcal{F}_{X|\mathcal{Q}}(x - c)$ for all $x \in \mathbb{R}$. Hence, we have

$$\widetilde{\rho}^{\mathrm{MA}_1}(X + c) = \rho\left(\bigvee_1 \mathcal{F}_{X+c|\mathcal{Q}}\right) = \rho\left(\bigvee_1 \mathcal{F}_{X|\mathcal{Q}}\right) + c = \widetilde{\rho}^{\mathrm{MA}_1}(X) + c,$$

which yields translation invariance of $\widetilde{\rho}^{\mathrm{MA}_1}$. For MA_2, using Proposition 1 again, we have

$$\pi_{\bigvee_2 \mathcal{F}_{X+c|\mathcal{Q}}}(x) = \sup_{Q\in\mathcal{Q}} \mathbb{E}^{Q}[(X + c - x)_+] = \pi_{\bigvee_2 \mathcal{F}_{X|\mathcal{Q}}}(x - c), \quad \forall x \in \mathbb{R},$$

which implies $\bigvee_2 \mathcal{F}_{X+c|\mathcal{Q}}(x) = \bigvee_2 \mathcal{F}_{X|\mathcal{Q}}(x - c)$ for all $x \in \mathbb{R}$. Hence, we have

$$\widetilde{\rho}^{\mathrm{MA}_2}(X + c) = \rho\left(\bigvee_2 \mathcal{F}_{X+c|\mathcal{Q}}\right) = \rho\left(\bigvee_2 \mathcal{F}_{X|\mathcal{Q}}\right) + c = \widetilde{\rho}^{\mathrm{MA}_2}(X) + c,$$

which shows that $\widetilde{\rho}^{\mathrm{MA}_2}$ is translation invariant. □

E Proofs for results in Sections 5 and omitted examples

E.1 Proofs

Proof of Proposition 3. We introduce some notation that defined in the proof of Theorem 2 (a). Let $g_F(\alpha) = \mathrm{ES}_\alpha(F)$ for $\alpha \in (0,1)$ and $F \in \mathcal{M}_1$. Define $\mathcal{G} = \{g : (0,1) \to \mathbb{R} | (1 - \alpha)g(\alpha)$ is concave for $\alpha \in (0,1)$, $\lim_{\alpha\to 1}(1-\alpha)g(\alpha) = 0\}$, and $\widehat{\rho} : \mathcal{G} \to \mathbb{R}$ as $\widehat{\rho}(g_F) = \rho(F)$. As shown in the proof of Theorem 2 (a), we know that ρ is $\preceq_2$-consistent. Hence, we have

$$\begin{aligned}\rho^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f}) &= \inf\left\{\rho(G) : G \in \mathcal{M}_1,\ \sup_{F\in\mathcal{F}} \widetilde{\mathrm{ES}}^F_\alpha(f(\mathbf{a},\mathbf{X})) \leqslant g_G(\alpha),\quad \forall \alpha \in (0,1)\right\} \\ &= \inf\left\{\widehat{\rho}(g) : g \in \mathcal{G},\ \sup_{F\in\mathcal{F}} \widetilde{\mathrm{ES}}^F_\alpha(f(\mathbf{a},\mathbf{X})) \leqslant g(\alpha),\quad \forall \alpha \in (0,1)\right\} \\ &= \inf_{g\in\mathcal{G}} \left\{\widehat{\rho}(g) + \Theta(\mathbf{a}, g)\right\},\end{aligned}$$

where $\Theta(\mathbf{a}, g) = 0$ if $\sup_{F\in F} \widetilde{\mathrm{ES}}^F_\alpha(f(\mathbf{a},\mathbf{X})) \leqslant g(\alpha)$ for all $\alpha \in (0,1)$, and $\Theta(\mathbf{a}, g) = \infty$ otherwise. The remained proof is similar to that of Theorem 2 (a) by noting that $f(\mathbf{a},\mathbf{x})$ is convex in $\mathbf{a}$. □

The following proposition shows that any law-invariant coherent risk measure on $\mathcal{M}_1$ can be approximated by risk measures in the form of (17). Recall that any law-invariant coherent risk measure that is finite on $\mathcal{M}_1$ admits a Kusuoka representation (see e.g., Shapiro (2013))

$$\rho(F) = \sup_{\mu\in\mathcal{P}_0} \int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha), \qquad F \in \mathcal{M}_1, \tag{EC.9}$$

where $\mathcal{P}_0$ is a subset of the set $\mathcal{P}$ of all probability measures on $[0,1]$.

Proposition EC.2. *Let $\rho : \mathcal{M}_1 \to \mathbb{R}$ be a law-invariant coherent risk measure. There exists a sequence of risk measures $(\rho_n)_{n\in\mathbb{N}}$ of the form (17) such that $\lim_{n\to\infty} \rho_n(F) = \rho(F)$ for any $F \in \mathcal{M}_1$. Moreover, if $\mathcal{P}_0$ in (EC.9) is finite, then for any set $\mathcal{G}$ of distributions supported within a common compact interval, uniform convergence holds: $\sup_{F\in\mathcal{G}} |\rho_n(F) - \rho(F)| \to 0$ as $n \to \infty$.*

Proof. Denote by $D_n = \{i/2^n : i = 0, \ldots, 2^n\}$, $n \in \mathbb{N}$, and $D = \bigcup_{n\in\mathbb{N}} D_n$. Note that D is countable and dense in $[0,1]$. For $n \in \mathbb{N}$, denote by $\mathcal{P}_n$ the finite set of convex combinations of elements of $\{\delta_j : j \in D_n\}$ with weights in D_n, that is,

$$\mathcal{P}_n = \left\{\sum_{i=0}^{2^n} \alpha_i \delta_{i/2^n} :\ \alpha_i \in D_n,\ i = 0, \ldots, 2^n,\ \sum_{i=0}^{2^n} \alpha_i = 1\right\}.$$

Let $\mathcal{P}_0$ be given by (EC.9), which represents ρ. For each $\mu \in \mathcal{P}_0$, define $\mu_n = \sum_{i=0}^{2^n-1} \alpha_{n,i}\delta_{i/2^n} \in \mathcal{P}_n$, $n \in \mathbb{N}$, where $\alpha_{n,i} = \beta_{n,i} - \beta_{n,i-1}$ and

$$\beta_{n,i} = 2^{-n}\left\lceil 2^n \mu\left(\left[0, \frac{i+1}{2^n}\right]\right)\right\rceil, \quad i = 0, \ldots, 2^n - 1; \quad \beta_{n,-1} = 0.$$

Here, $\lceil x\rceil$ is the smallest integer dominating x. This construction guarantees that $\mu_n \preceq_1 \mu$ and μ_n converges to μ in distribution as $n \to \infty$. Therefore, for $F \in \mathcal{M}_1$, we have that $\alpha \mapsto \mathrm{ES}_\alpha(F)$ is continuous and bounded from below, and hence,

$$\liminf_{n\to\infty}\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu_n(\alpha) \geqslant \int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha).$$

Let $\mathbb{W}_n = \bigcup_{\mu\in\mathcal{P}_0}\{\mu_n\}$ and

$$\rho_n(F) = \sup_{\nu\in\mathbb{W}_n}\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\nu(\alpha), \quad F\in\mathcal{M}_1.$$

Note that $\mathbb{W}_n \subseteq \mathcal{P}_n$ is a finite set, and hence ρ_n is a risk measure of the form (17). For each $\mu\in\mathcal{P}_0$,

$$\liminf_{n\to\infty}\rho_n(F) \geqslant \liminf_{n\to\infty}\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu_n(\alpha) \geqslant \int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha).$$

Therefore, we have for $F\in\mathcal{M}_1$,

$$\liminf_{n\to\infty}\rho_n(F) \geqslant \sup_{\mu\in\mathcal{P}_0}\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha) = \rho(F).$$

On the other hand, by $\mu_n \preceq_1 \mu$, we have $\rho_n(F)\leqslant\rho(F)$ for all $n\in\mathbb{N}$, and thus, $\lim_{n\to\infty}\rho_n(F) = \rho(F)$. This completes the proof of the first statement.

Next, we show uniform convergence on $\mathcal{G}$, assuming that $\mathcal{P}_0$ is finite. Note that

$$\begin{aligned}\sup_{F\in\mathcal{G}}|\rho_n(F)-\rho(F)| &= \sup_{F\in\mathcal{G}}\left|\sup_{\mu\in\mathcal{P}_0}\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu_n(\alpha) - \sup_{\mu\in\mathcal{P}_0}\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha)\right| \\ &\leqslant \sup_{\mu\in\mathcal{P}_0}\sup_{F\in\mathcal{G}}\left|\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu_n(\alpha) - \int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha)\right|.\end{aligned}$$

Recall that ρ and ρ_n for $n\in\mathbb{N}$ satisfy translation invariance and positive homogeneity. To see $\lim_{n\to\infty}\sup_{F\in\mathcal{G}}|\rho_n(F)-\rho(F)| = 0$, it suffices to verify that for $\mu\in\mathcal{P}_0$

$$\lim_{n\to\infty}\sup_{F\in\mathcal{M}[0,1]}\left|\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu_n(\alpha) - \int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha)\right| = 0, \tag{EC.10}$$

where $\mathcal{M}[0,1]$ is the set of all distributions with support in $[0,1]$. Below let us prove (EC.10). Define, for $s\in[0,1]$,

$$G(s)=\mu([0,s]),\quad G_n(s)=\mu_n([0,s]),\quad h(s)=\int_0^s \frac{1}{1-\alpha}\mathrm{d}G(\alpha),\quad \text{and}\quad h_n(s)=\int_0^s \frac{1}{1-\alpha}\mathrm{d}G_n(\alpha).$$

Applying integration by parts, we have

$$h_n(s) - h(s) = \frac{G_n(s)}{1-s} - \int_0^s \frac{G_n(\alpha)}{(1-\alpha)^2}\mathrm{d}\alpha - \frac{G(s)}{1-s} + \int_0^s \frac{G(\alpha)}{(1-\alpha)^2}\mathrm{d}\alpha.$$

Let A be the set of all continuity points of G. Since $\mu_n \to \mu$ in distribution, we have $G_n(\alpha) \to G(\alpha)$ for all $\alpha \in A$, and this also implies $\int_0^s G_n(\alpha)/(1-\alpha)^2\mathrm{d}\alpha \to \int_0^s G(\alpha)/(1-\alpha)^2\mathrm{d}\alpha$ for all $s \in [0,1)$. Hence, we have

$$h_n(s) - h(s) \to 0 \quad \text{for } s \in A. \tag{EC.11}$$

Using the relation that $(1-\alpha)\mathrm{ES}_\alpha(F) = \int_\alpha^1 \mathrm{VaR}_s(F)\mathrm{d}s$, we have

$$\begin{aligned}
&\sup_{F\in\mathcal{M}[0,1]} \left|\int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu_n(\alpha) - \int_0^1 \mathrm{ES}_\alpha(F)\mathrm{d}\mu(\alpha)\right| \\
&= \sup_{F\in\mathcal{M}[0,1]} \left|\int_0^1 \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s(F)\mathrm{d}s\mathrm{d}\mu_n(\alpha) - \int_0^1 \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_s(F)\mathrm{d}s\mathrm{d}\mu(\alpha)\right| \\
&= \sup_{F\in\mathcal{M}[0,1]} \left|\int_0^1 \mathrm{VaR}_s(F)h_n(s)\mathrm{d}s - \int_0^1 \mathrm{VaR}_s(F)h(s)\mathrm{d}s\right| \\
&\leqslant \sup_{F\in\mathcal{M}[0,1]} \int_0^1 \left|(h_n(s)-h(s))\mathrm{VaR}_s(F)\right|\mathrm{d}s \leqslant \int_0^1 |h_n(s)-h(s)|\mathrm{d}s,
\end{aligned}$$

where we have used Fubini's theorem in the second step. By (EC.11), we know that $h_n \to h$ almost surely on $[0,1]$. On the other hand, it follows from Theorem 4 of Huang and Wang (2024) that $h(1) < \infty$. Note that $\mu_n \preceq_1 \mu$ and h_n is increasing on $[0,1]$. For any $s \in [0,1]$ and $n \in \mathbb{N}$, we have

$$0 \leqslant h_n(s) \leqslant h_n(1) \leqslant h(1) < \infty,$$

and this implies that $\{h_n\}_{n\in\mathbb{N}}$ is a bounded sequence. Therefore, we conclude that $\lim_{n\to\infty}\int_0^1 |h_n(s) - h(s)|\mathrm{d}s = 0$, and (EC.10) holds. This completes the proof. □

Proof of Proposition 3. Note that $\min_{\mathbf{a}\in A}\rho^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f})$ is equivalent to

$$\min_{\mathbf{a}\in A}\sup_{w\in\mathbb{W}}\sum_{j=1}^{n^w} p_j^w \mathrm{ES}_{\alpha_j^w}\left(\bigvee_2 \mathcal{F}_{\mathbf{a},f}\right), \tag{EC.12}$$

where $\mathcal{F}_{\mathbf{a},f} = \{F_{f(\mathbf{a},\mathbf{X})} : F_{\mathbf{X}} \in \mathcal{F}\}$. It holds that

$$\mathrm{ES}_{\alpha_j^w}\left(\bigvee_2 \mathcal{F}_{\mathbf{a},f}\right) = \inf_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\alpha_j^w}\pi_{\vee_2\mathcal{F}_{\mathbf{a},f}}(x)\right\} = \inf_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\alpha_j^w}\sup_{F\in\mathcal{F}}\mathbb{E}^F\left[(f(\mathbf{a},\mathbf{X})-x)_+\right]\right\},$$

where the first and the second steps follow from (12) and Proposition 1, respectively. Substituting the above equation into (EC.12) yields the following equivalent problem

$$\begin{aligned}
&\min_{\mathbf{a}\in A, h_j^w\in\mathbb{R}} \sup_{w\in\mathbb{W}} \sum_{j=1}^{n^w} p_j^w h_j^w \\
&\text{s.t.}\ \inf_{x\in\mathbb{R}}\left\{x + \frac{1}{1-\alpha_j^w}\sup_{F\in\mathcal{F}}\mathbb{E}^F[(f(\mathbf{a},\mathbf{X})-x)_+]\right\} \leqslant h_j^w, \quad j\in[n^w],\ w\in\mathbb{W}.
\end{aligned}$$

This is again equivalent to Problem (18) which completes the proof. □

Proof of (19). Note that $\min_{\mathbf{a}\in A}\rho^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f})$ is equivalent to

$$\min_{\mathbf{a}\in A}\ \sup_{i\in[n]}\sup_{w\in\mathbb{W}}\sum_{j=1}^{n^w} p_j^w \widetilde{\mathrm{ES}}_{\alpha_j^w}^{F_i}(f(\mathbf{a},\mathbf{X})),$$

that is,

$$\begin{aligned}
&\min_{\mathbf{a}\in A,\, h\in\mathbb{R}}\ h \\
&\quad\text{s.t.}\ \sum_{j=1}^{n^w} p_j^w \widetilde{\mathrm{ES}}_{\alpha_j^w}^{F_i}(f(\mathbf{a},\mathbf{X})) \leqslant h, \quad i\in[n],\ w\in\mathbb{W}.
\end{aligned}$$

This is again equivalent to

$$\begin{aligned}
&\min_{\mathbf{a}\in A, h, h_{i,j}^w\in\mathbb{R}}\ h \\
&\quad\text{s.t.}\ \sum_{j=1}^{n^w} p_j^w h_{i,j}^w \leqslant h, \quad i\in[n],\ w\in\mathbb{W} \\
&\qquad\ \widetilde{\mathrm{ES}}_{\alpha_j^w}^{F_i}(f(\mathbf{a},\mathbf{X})) \leqslant h_{i,j}^w,\ i\in[n],\ j\in[n^w],\ w\in\mathbb{W}.
\end{aligned}$$

Substituting the representation $\widetilde{\mathrm{ES}}_{\alpha}^{F}(f(\mathbf{a},\mathbf{X})) = \inf_{x\in\mathbb{R}}\{x + \frac{1}{1-\alpha}\mathbb{E}^F[(f(\mathbf{a},\mathbf{X})-x)_+]\}$ into the above problem yields (19). □

E.2 Examples

First, we propose the following example to demonstrate that the convexity of WR and MA_2 robust optimization problems may not always coincide.

Example 3. Let $A \subseteq \mathbb{R}_+^2$ be convex. Define $f : A \times \mathbb{R} \to \mathbb{R}$ as $f(\mathbf{a}, x) = (a_1 \vee a_2)x$, where $a_1 \vee a_2 = \max\{a_1, a_2\}$. Denote by $\mathcal{F} = \{F_1, F_2\}$ with $F_1 = \delta_{-1}$ and $F_2 = 2\delta_{-9}/3 + \delta_9/3$ where δ_t represents the point-mass at $t \in \mathbb{R}$. For $\alpha \in [1/6, 1/3)$, we consider the WR and MA_2 robust optimization problems:

$$\min_{\mathbf{a}\in A} \mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f}) \quad \text{and} \quad \min_{\mathbf{a}\in A} \mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f}). \tag{EC.13}$$

For WR robust optimization, we have

$$\begin{aligned}\mathrm{ES}_\alpha^{\mathrm{WR}}(\mathcal{F}_{\mathbf{a},f}) &= \max_{F\in\mathcal{F}} \widetilde{\mathrm{ES}}_\alpha^F(f(\mathbf{a}, X)) = (\mathrm{ES}_\alpha(F_1) \vee \mathrm{ES}_\alpha(F_2))(a_1 \vee a_2)\\ &= \left((-1) \vee \left(\frac{6}{1-\alpha} - 9\right)\right)(a_1 \vee a_2).\end{aligned}$$

Note that $6/(1-\alpha) - 9 < 0$ as $\alpha \in [1/6, 1/3)$. We have WR robust optimization in (EC.13) is not a convex optimization problem. For MA_2 robust optimization, one can check that

$$\bigvee_2 \mathcal{F}_{\mathbf{a},f} = \frac{2}{3}\delta_{-6(a_1\vee a_2)} + \frac{1}{3}\delta_{9(a_1\vee a_2)}, \quad \forall \mathbf{a} \in A.$$

Hence,

$$\mathrm{ES}_\alpha^{\mathrm{MA}_2}(\mathcal{F}_{\mathbf{a},f}) = \mathrm{ES}_\alpha\left(\bigvee_2 \mathcal{F}_{\mathbf{a},f}\right) = \left(\frac{5}{1-\alpha} - 6\right)(a_1 \vee a_2).$$

Note that $5/(1-\alpha) - 6 \geqslant 0$ as $\alpha \in [1/6, 1/3)$. We have MA_2 robust optimization in (EC.13) is a convex optimization problem. If we let $f(\mathbf{a}, x) = (a_1 \wedge a_2) \cdot x$, where $a_1 \wedge a_2 = \min\{a_1, a_2\}$, then it follows the similar arguments previously that the WR optimization problem is convex, and the MA_2 optimization problem is not convex.

Next, we study the tractability of the MA_2 robust optimization under two common convex uncertainty sets, namely box uncertainty and ellipsoid uncertainty, that are specifically associated with discrete cdfs.

Example 4. Suppose that the sample space is $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$, and consider the uncertainty set

$$\mathcal{F}_\Theta = \left\{ \sum_{i=1}^{N} \theta_i \delta_{\mathbf{x}_i} : \boldsymbol{\theta} \in \Theta \right\},$$

where Θ is a subset of Δ_N. We have that for any fixed $\mathbf{a} \in A$ and $x \in \mathbb{R}$,

$$\sup_{F \in \mathcal{F}_\Theta} \mathbb{E}^F[(f(\mathbf{a}, \mathbf{X}) - x)_+] = \sup_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^{N} \theta_i (f(\mathbf{a}, \mathbf{x}_i) - x)_+.$$

This implies that Problem (18) can be equivalently reformed as the following problem with variables $\mathbf{a} \in A$ and $h \in \mathbb{R}$; $\mathbf{h}^w, \mathbf{x}^w \in \mathbb{R}^{n^w}$ for $w \in \mathbb{W}$; $\mathbf{u}^{j,w} \in \mathbb{R}^N$ for $j \in n^w$ and $w \in \mathbb{W}$:

$$\begin{aligned} \min \ & h \\ \text{s.t.} \ & (\mathbf{p}^w)^\top \mathbf{h}^w \leqslant h, \quad w \in \mathbb{W} \\ & x_j^w + \frac{1}{1-\alpha_j^w} \sup_{\boldsymbol{\theta} \in \Theta} \boldsymbol{\theta}^\top \mathbf{u}^{j,w} \leqslant h_j^w, \quad j \in [n^w], \ w \in \mathbb{W} \\ & u_i^{j,w} \geqslant f(\mathbf{a}, \mathbf{x}_i) - x_j^w, \quad i \in [N], \ j \in [n^w], \ w \in \mathbb{W} \\ & u_i^{j,w} \geqslant 0, \quad i \in [N], \ j \in [n^w], \ w \in \mathbb{W}. \end{aligned} \tag{EC.14}$$

When Θ is chosen as a box uncertainty or an ellipsoid uncertainty, we can further reform the above problem as follows.

(i) **Box uncertainty.** We set Θ as a box, i.e.,

$$\Theta = \Theta_B = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \boldsymbol{\theta}^0 + \boldsymbol{\eta}, \ \mathbf{e}^\top \boldsymbol{\eta} = 0, \ \underline{\boldsymbol{\eta}} \leqslant \boldsymbol{\eta} \leqslant \overline{\boldsymbol{\eta}}\},$$

where $\boldsymbol{\theta}^0$ is a nominal distribution, $\mathbf{e}$ is the vector of ones, and $\underline{\boldsymbol{\eta}}$ and $\overline{\boldsymbol{\eta}}$ are given vectors. The constraints of $\boldsymbol{\eta}$ in the above set ensure that $\Theta_B \subseteq \Delta_N$. Using a similar argument of

Section 2.2.1 in Zhu and Fukushima (2009), Problem (EC.14) is equivalent to

$$\begin{aligned}
\min\ & h && \text{(EC.15)}\\
\text{s.t.}\ & (\mathbf{p}^w)^\top \mathbf{h}^w \leqslant h, \quad w \in \mathbb{W}\\
& x_j^w + \frac{1}{1-\alpha_j^w}(\boldsymbol{\theta}^0)^\top \mathbf{u}^{j,w} + \frac{1}{1-\alpha_j^w}(\overline{\boldsymbol{\eta}}^\top \boldsymbol{\xi}^{j,w} + \underline{\boldsymbol{\eta}}^\top \boldsymbol{\gamma}^{j,w}) \leqslant h_j^w, \quad j \in [n^w],\ w \in \mathbb{W}\\
& \mathbf{e} z^{j,w} + \boldsymbol{\xi}^{j,w} + \boldsymbol{\gamma}^{j,w} = 0,\ \boldsymbol{\xi}^{j,w} \geqslant 0,\ \boldsymbol{\gamma}^{j,w} \leqslant 0, \quad j \in [n^w],\ w \in \mathbb{W}\\
& u_i^{j,w} \geqslant f(\mathbf{a}, \mathbf{x}_i) - x_j^w, \quad i \in [N],\ j \in [n^w],\ w \in \mathbb{W}\\
& u_i^{j,w} \geqslant 0, \quad i[N],\ j \in [n^w],\ w \in \mathbb{W},
\end{aligned}$$

where variables are $\mathbf{a} \in A$ and $h \in \mathbb{R}$; $\mathbf{h}^w, \mathbf{x}^w \in \mathbb{R}^{n^w}$ for $w \in \mathbb{W}$; $z^{j,w} \in \mathbb{R}$ and $\mathbf{u}^{j,w}, \boldsymbol{\xi}^{j,w}, \boldsymbol{\gamma}^{j,w} \in \mathbb{R}^N$ for $j \in [n^w]$ and $w \in \mathbb{W}$. If $f(\mathbf{a}, \mathbf{x})$ is convex in terms of $\mathbf{a}$ and A is convex, then Problem (EC.15) is a convex optimization problem that is consistent with Proposition 3. Specifically, when $f(\mathbf{a}, \mathbf{x})$ is linear in $\mathbf{a}$ and A is a convex polyhedron, the optimization problem simplifies to a linear program.

(ii) **Ellipsoid uncertainty.** We set Θ as an ellipsoid, i.e.,

$$\Theta = \Theta_E = \{\boldsymbol{\theta} : \boldsymbol{\theta} = \boldsymbol{\theta}^0 + C\boldsymbol{\eta},\ \mathbf{e}^\top C\boldsymbol{\eta} = 0,\ \boldsymbol{\theta}^0 + C\boldsymbol{\eta} \geqslant 0,\ \|\boldsymbol{\eta}\|_2 \leqslant 1\},$$

where $\|\boldsymbol{\eta}\|_2 = \sqrt{\boldsymbol{\eta}^\top \boldsymbol{\eta}}$, $\boldsymbol{\theta}^0$ is a nominal distribution which is also the center of the ellipsoid, and $C \in \mathbb{R}^{N\times N}$ is the scaling matrix of the ellipsoid. The constraints of $\boldsymbol{\eta}$ in the above set ensure that $\Theta_E \subseteq \Delta_N$. The dual problem of

$$\sup_{\boldsymbol{\eta}\in\mathbb{R}^N} \{\mathbf{u}^\top A\boldsymbol{\eta} : \mathbf{e}^\top C\boldsymbol{\eta} = 0,\ \boldsymbol{\theta}^0 + C\boldsymbol{\eta} \geqslant 0,\ \|\boldsymbol{\eta}\|_2 \leqslant 1\}$$

is an SOCP, which has the form:

$$\inf_{(\boldsymbol{\xi},\boldsymbol{\gamma},\zeta,z)\in\mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}\times\mathbb{R}} \{\zeta + (\boldsymbol{\theta}^0)^\top \boldsymbol{\gamma} : -\boldsymbol{\xi} - C^\top \boldsymbol{\gamma} + C^\top \mathbf{e} z = C^\top \mathbf{u},\ \|\boldsymbol{\xi}\|_2 \leqslant \zeta,\ \boldsymbol{\gamma} \geqslant 0\}.$$

The strong duality between above two problems holds under some mild condition such as the Slater's condition. In this case, using a similar argument of Section 2.2.2 in Zhu and

Fukushima (2009), Problem (EC.14) can be equivalently reformed as

$$\begin{aligned}
\min\ & h && \text{(EC.16)}\\
\text{s.t. } & (\mathbf{p}^w)^\top \mathbf{h}^w \leqslant h, \quad w \in \mathbb{W}\\
& x_j^w + \frac{1}{1-\alpha_j^w}(\boldsymbol{\theta}^0)^\top \mathbf{u}^{j,w} + \frac{1}{1-\alpha_j^w}(\zeta^{j,w} + (\boldsymbol{\theta}^0)^\top \boldsymbol{\gamma}^{j,w}) \leqslant h_j^w, \quad j \in [n^w],\ w \in \mathbb{W}\\
& -\boldsymbol{\xi}^{j,w} - C^\top \boldsymbol{\gamma}^{j,w} + C^\top \mathbf{e} z^{j,w} = C^\top \mathbf{u}^{j,w},\ \|\boldsymbol{\xi}^{j,w}\|_2 \geqslant \zeta^{j,w},\ \boldsymbol{\gamma}^{j,w} \leqslant 0, \quad j \in [n^w],\ w \in \mathbb{W}\\
& u_i^{j,w} \geqslant f(\mathbf{a}, \mathbf{x}_i) - x_j^w, \quad i \in [N],\ j \in [n^w], n^w,\ w \in \mathbb{W}\\
& u_i^{j,w} \geqslant 0, \quad i \in [N],\ j \in [n^w],\ w \in \mathbb{W},
\end{aligned}$$

where variables are $\mathbf{a} \in A$ and $h \in \mathbb{R}$; $\mathbf{h}^w, \mathbf{x}^w \in \mathbb{R}^{n^w}$ for $w \in \mathbb{W}$; $z^{j,w}, \zeta^{j,w} \in \mathbb{R}$ and $\mathbf{u}^{j,w}, \boldsymbol{\xi}^{j,w}, \boldsymbol{\gamma}^{j,w} \in \mathbb{R}^N$ for $j \in [n^w]$ and $w \in \mathbb{W}$. If $f(\mathbf{a}, \mathbf{x})$ is convex in $\mathbf{a}$ and A is a convex set, then Problem (EC.16) is a convex optimization problem that is consistent with Proposition 3. Additionally, if $f(\mathbf{a}, \mathbf{x})$ is linear in $\mathbf{a}$ and A is a convex polyhedron, the problem simplifies to an SOCP, which can be efficiently solved using interior-point methods.

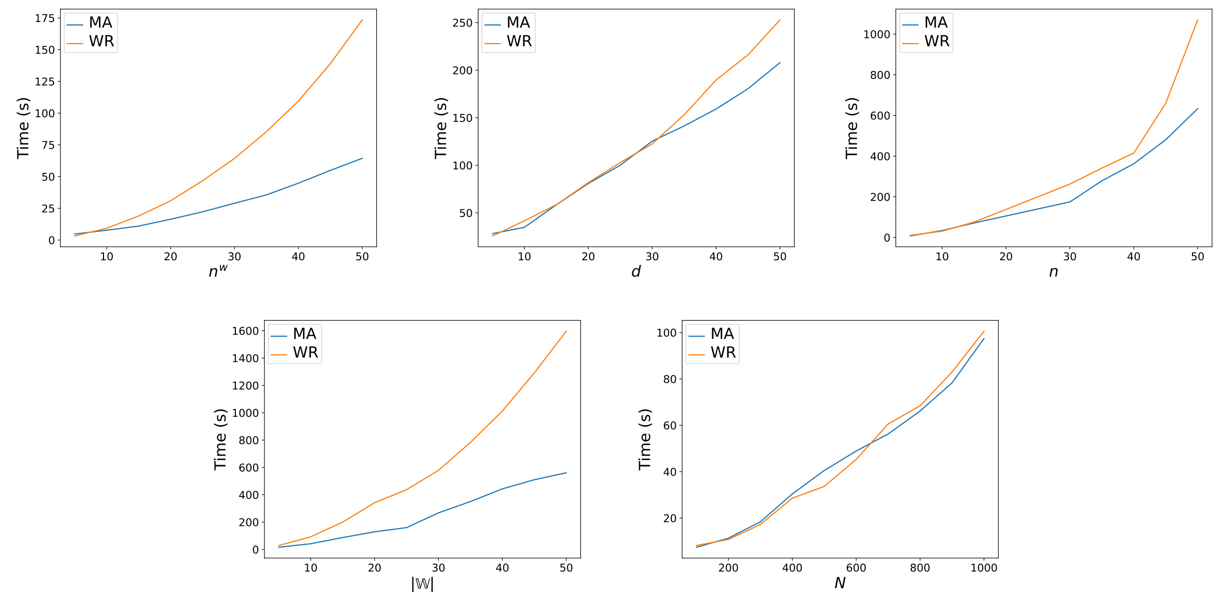
To end this section, we compare the tractability of (19) and (20). Set

$$f(\mathbf{a}, \mathbf{x}) = \sum_{i=1}^{d} [\beta_i (a_i - x_i)_+ + \eta_i (x_i - a_i)_+],$$

where $\beta_i, \eta_i \in \mathbb{R}_+$. This is the loss function of the newsvendor problem.[17] For $i \in [n]$, we take F_i as the empirical distribution of N sample points from the distribution $\mathrm{N}(\boldsymbol{\mu}_i, \Sigma_i)$; this can be seen as an approximation of the corresponding normal distributions via Monte Carlo simulation. We generate the parameters by $(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i) \overset{\text{i.i.d.}}{\sim} \mathrm{N}_{2d}(0, I_{2d})$ and $\Sigma_i = \mathrm{diag}(\exp(\boldsymbol{\sigma}_i))$ for $i \in [n]$, where I_d is the $d \times d$ identity matrix and exp is applied component-wise. Figure EC.1 presents the computation times of two approaches where we use the `cvxpy` library with `MOSEK` as the chosen solver. The corresponding parameters are summarized below. We set $\beta_i = i$ and $\eta_i = d - i + 1$ for $i \in [d]$. We assume that the values of n^w are identical for all $w \in \mathbb{W}$, and p_j^w, $j \in [n^w]$ are uniformly generated from the standard n^w-simplex, and $\alpha_j^w = (2j-1)/(2n^w)$ for $j \in [n^w]$. Each panel in Figure EC.1 illustrates the performance as one parameter varies, while the other parameters are held constant at $(n^w, d, n, |\mathbb{W}|, N) = (10, 3, 3, 3, 100)$. As shown in Figure EC.1, the computation times of the two approaches appear to be similar for most values of d, n and N. In the case of large n^w or large $|\mathbb{W}|$, the MA approach is visibly faster. This is consistent with our intuition that MA

[17]See Fu et al. (2022) for a related distributionally robust newsvendor optimization based on ambiguity set constructed based on $\preceq_1$.

Figure EC.1: Computation times of MA and WR methods



Note: d is the dimension of the random vector; n is the number of cdfs in $\mathcal{F}$; $|\mathbb{W}|$ is the cardinality of $\mathbb{W}$; N is the sample size drawn from each $F_i \in \mathcal{F}$. We assume that n^w is identical for each $w \in \mathbb{W}$.

may work better when the risk measure itself is more complicated.

F Proofs for results in Section 6 and omitted figures

F.1 Proofs

Proof of Theorem 4. Statement (a) can be directly obtained by applying Proposition 4 (i) of Liu et al. (2022). To see (b), if $p = 1$, one can check that $\sup_{F\in\mathcal{F}_{1,\varepsilon}(F_0)} \pi_F(x) = \pi_{F_0}(x) + \varepsilon$ for all $x \in \mathbb{R}$. There does not exist $G \in \mathcal{M}_1$ such that $\pi_G = \sup_{F\in\mathcal{F}_{1,\varepsilon}(F_0)} \pi_F$ since $\lim_{x\to\infty} \sup_{F\in\mathcal{F}_{1,\varepsilon}(F_0)} \pi_F(x) = \varepsilon > 0$. Hence, the set $\mathcal{F}_{1,\varepsilon}(F_0)$ is not $\preceq_2$-bounded. For $p > 1$, Since the Wasserstein ball is convex, it follows from Theorem 1 that $\sup_{F\in\mathcal{F}_{p,\varepsilon}(F_0)} \mathrm{ES}_\alpha(F) = \mathrm{ES}_\alpha(F^2_{p,\varepsilon|F_0})$. By Proposition 4 (ii) of Liu et al. (2022), we have $\sup_{F\in\mathcal{F}_{p,\varepsilon}(F_0)} \mathrm{ES}_\alpha(F) = \mathrm{ES}_\alpha(F_0) + (1-\alpha)^{-1/p}\varepsilon$ for $\alpha \in (0,1)$. Therefore, one can obtain

$$\int_\alpha^1 \mathrm{VaR}_s\left(F^2_{p,\varepsilon|F_0}\right) \mathrm{d}s = \int_\alpha^1 \mathrm{VaR}_s(F_0)\mathrm{d}s + (1-\alpha)^{1-\frac{1}{p}}\varepsilon, \quad \alpha \in (0,1).$$

Take the derivative on the left and right sides of the above formula for α, we have

$$\mathrm{VaR}_\alpha\left(F^2_{p,\varepsilon|F_0}\right) = \mathrm{VaR}_\alpha(F_0) + \left(1 - \frac{1}{p}\right)(1-\alpha)^{-\frac{1}{p}}\varepsilon.$$

Hence, we complete the proof. $\square$

The following remark is related to the robust distributions in Theorem 4.

Remark EC.2. In this remark, we collect some observations related to the robust distributions $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ obtained in Theorem 4.

(i) The order $F^2_{p,\varepsilon|F_0} \preceq_1 F^1_{p,\varepsilon|F_0}$ holds since $(F^1_{p,\varepsilon|F_0})^{-1}(\alpha) \geqslant (F^2_{p,\varepsilon|F_0})^{-1}(\alpha)$ for all $\alpha \in (0,1)$.

(ii) Both $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ are increasing in ε with respect to $\preceq_1$.

(iii) The left-hand side of equation (23) is increasing in p. Hence, a larger value of p leads to a smaller cdf $F^1_{p,\varepsilon|F_0}$ with respect to $\preceq_1$.

(iv) The left quantile functions $(F^1_{p,\varepsilon|F_0})^{-1}$ and $(F^2_{p,\varepsilon|F_0})^{-1}$ has the same limit $F_0^{-1} + \varepsilon$ as $p \to \infty$.

(v) None of $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ is in any Wasserstein ball of the form (22) since $W_p(F^1_{p,\varepsilon|F_0}, F_0) = W_p(F^2_{p,\varepsilon|F_0}, F_0) = \infty$. This is not surprising, as $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ dominate every element in the Wasserstein ball and their quantile functions are of a different shape in general.

Proof of Theorem 5. For two random vectors $\mathbf{X}$ and $\mathbf{Y}$ of the same dimension, define $\mathcal{L}_{a,p}$ as

$$\mathcal{L}_{a,p}(\mathbf{X},\mathbf{Y})^p = \mathbb{E}\left[\|\mathbf{X}-\mathbf{Y}\|_a^p\right].$$

For any $F \in \mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}})$, by definition, there exists $\mathbf{Z}$ with $F_{\mathbf{Z}} \in \mathcal{F}^d_{a,p,\varepsilon}(F_{\mathbf{X}})$ and $F = F_{\mathbf{w}^\top \mathbf{Z}}$. We can verify that

$$\begin{aligned}
W_p(F_{\mathbf{w}^\top\mathbf{X}}, F_{\mathbf{w}^\top\mathbf{Z}}) &= \inf_{X \overset{\mathrm{d}}{=} \mathbf{w}^\top\mathbf{X},\ Z \overset{\mathrm{d}}{=} \mathbf{w}^\top\mathbf{Z}} (\mathbb{E}[|X-Z|^p])^{1/p} \\
&= \inf_{\mathbf{w}^\top\mathbf{X}' \overset{\mathrm{d}}{=} \mathbf{w}^\top\mathbf{X},\ \mathbf{w}^\top\mathbf{Z}' \overset{\mathrm{d}}{=} \mathbf{w}^\top\mathbf{Z}} (\mathbb{E}[|\mathbf{w}^\top\mathbf{X}' - \mathbf{w}^\top\mathbf{Z}'|^p])^{1/p} \\
&\leqslant \inf_{\mathbf{w}^\top\mathbf{X}' \overset{\mathrm{d}}{=} \mathbf{w}^\top\mathbf{X},\ \mathbf{w}^\top\mathbf{Z}' \overset{\mathrm{d}}{=} \mathbf{w}^\top\mathbf{Z}} \|\mathbf{w}\|_b \mathcal{L}_{a,p}(\mathbf{X}',\mathbf{Z}') \\
&\leqslant \inf_{\mathbf{X}' \overset{\mathrm{d}}{=} \mathbf{X},\ \mathbf{Z}' \overset{\mathrm{d}}{=} \mathbf{Z}} \|\mathbf{w}\|_b \mathcal{L}_{a,p}(\mathbf{X}',\mathbf{Z}') = \|\mathbf{w}\|_b W^d_{a,p}(F_{\mathbf{X}}, F_{\mathbf{Z}}) \leqslant \varepsilon\|\mathbf{w}\|_b,
\end{aligned}$$

where the infima are taken over (X,Z) or $(\mathbf{X}',\mathbf{Z}')$, b satisfies $1/a + 1/b = 1$, and the first inequality follows from the Hölder inequality. Hence, $\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}}) \subseteq \mathcal{F}_{p,\varepsilon\|\mathbf{w}\|_b}(F_{\mathbf{w}^\top\mathbf{X}})$.

We next show the opposite direction of the set inclusion. For any $F \in \mathcal{F}_{p,\varepsilon\|\mathbf{w}\|_b}(F_{\mathbf{w}^\top\mathbf{X}})$, since the set $\{Y : (\mathbb{E}[|Y - \mathbf{w}^\top\mathbf{X}|^p])^{1/p} \leqslant \varepsilon\|\mathbf{w}\|_b\}$ is closed, there exists Z such that $F_Z = F$ and $(\mathbb{E}[|Z - \mathbf{w}^\top\mathbf{X}|^p])^{1/p} \leqslant \varepsilon\|\mathbf{w}\|_b$. Below we consider the case that $a > 1$. Denote by $Y = Z - \mathbf{w}^\top\mathbf{X}$ and $\mathbf{Z} = \mathbf{X} + (\mathbf{w}^{b/a}Y)/(\mathbf{w}^\top\mathbf{w}^{b/a})$, where $\mathbf{w}^{b/a} = (\mathrm{sign}(w_1)|w_1|^{b/a}, \ldots, \mathrm{sign}(w_d)|w_d|^{b/a})$ and sign:

$\mathbb{R} \to \{-1, 1\}$ is the sign function. We have $\mathbb{E}[|Y|^p] \leqslant \|\mathbf{w}\|_b^p \varepsilon^p$. Moreover, noting that $\mathbf{w}^\top \mathbf{w}^{b/a} = \sum_{i=1}^d |w_i|^{1+b/a} = \sum_{i=1}^d |w_i|^b = \|\mathbf{w}\|_b^b$, we have

$$\left(W_{a,p}^d(F_{\mathbf{X}}, F_{\mathbf{Z}})\right)^p \leqslant \mathbb{E}[\|\mathbf{Z} - \mathbf{X}\|_a^p] = \mathbb{E}\left[\left\|\frac{\mathbf{w}^{b/a}}{\mathbf{w}^\top \mathbf{w}^{b/a}} Y\right\|_a^p\right]$$

$$= \frac{\left\|\mathbf{w}^{b/a}\right\|_a^p \mathbb{E}[|Y|^p]}{\|\mathbf{w}\|_b^{pb}} = \frac{\|\mathbf{w}\|_b^{pb/a} \mathbb{E}[|Y|^p]}{\|\mathbf{w}\|_b^{pb}} = \frac{\mathbb{E}[|Y|^p]}{\|\mathbf{w}\|_b^p} \leqslant \varepsilon^p.$$

where the forth equality is due to $\left\|\mathbf{w}^{b/a}\right\|_a = \left(\sum_{i=1}^d |w_i|^b\right)^{1/a} = \|\mathbf{w}\|_b^{b/a}$ and the last equality comes from $b(1 - 1/a) = 1$. Hence, $F_{\mathbf{Z}} \in \mathcal{F}_{a,p,\varepsilon}^d(F_{\mathbf{X}})$, which implies $F_{\mathbf{w}^\top \mathbf{Z}} \in \mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}})$. Noting that $\mathbf{w}^\top \mathbf{Z} = \mathbf{w}^\top \mathbf{X} + Y = Z$, we obtain $F \in \mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}})$. This implies $\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}}) \supseteq \mathcal{F}_{p,\|\mathbf{w}\|_b \varepsilon}(F_{\mathbf{w}^\top \mathbf{X}})$. Hence, we conclude that $\mathcal{F}_{\mathbf{w},a,p,\varepsilon}(F_{\mathbf{X}}) = \mathcal{F}_{p,\|\mathbf{w}\|_b \varepsilon}(F_{\mathbf{w}^\top \mathbf{X}})$ for $a > 1$. If $a = 1$, then we let $i_0 \in [d]$ be such that $|w_{i_0}| = \|\mathbf{w}\|_\infty$, and denote by $Y = Z - \mathbf{w}^\top \mathbf{X}$ and $\mathbf{Z} = \mathbf{X} + \mathrm{sign}(w_{i_0}) \mathbf{e}_{i_0} Y / \|\mathbf{w}\|_\infty$, where $\mathbf{e}_{i_0}$ is a vector with its i_0-th element equal to 1 and all other elements equal to 0. Similar to the previous proof of the case that $a > 1$, the desired result holds. □

F.2 Omitted figures from Section 6

We present a few figures omitted from Section 6. Figures EC.2 and EC.3 are related to Section 6.1 for the Wasserstein uncertainty set $\mathcal{F}_{k,\varepsilon}(F_0)$ with $k = 2$, $\varepsilon = 0.1$, where the benchmark distribution F_0 is the standard normal distribution function. Figure EC.2 shows the left quantile functions of the supremum. In Figure EC.3, we obtain the WR and MA robust risk values and the risk measure is chosen as ES_α or PD_k.

Figure EC.4 shows the curves of robust risk evaluation via the WR and MA approaches for the mean-variance uncertainty set $\mathcal{F}_{0,1}$ in Section 6.3. The risk measure is chosen as ES_α, $\mathrm{RVaR}_{\alpha,\beta}$, PD_k or ex_α.

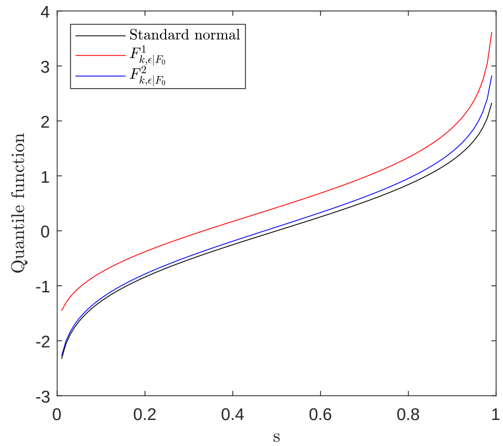
G Proofs and generalizations for results in Section 7

We first recall that, by Proposition 2, $\preceq_1$-cEMA and $\preceq_2$-cEMA imply the properties $\preceq_1$-consistency and $\preceq_2$-consistency, respectively, and $\preceq_i$-cEMA $(i = 1, 2)$ is equivalent to

$$\rho\left(\bigvee_i \{F_1, \dots, F_n\}\right) = \sup\left\{\rho\left(\sum_{i=1}^n \lambda_i F_i\right) : \boldsymbol{\lambda} \in \Delta_n\right\}$$

for all $F_1, \dots, F_n \in \mathcal{M}$ and $n \geqslant 1$.

Figure EC.2: The supremum of $\mathcal{F}_{2,0.1}(F_0)$ with $F_0 \sim N(0,1)$.



G.1 A generalization of Theorem 6 (a) and related results

The following theorem is a generalized version of Theorem 6 (a) to the domain $\mathcal{M}_p$, $p \in [0, \infty)$.

Theorem EC.1. *Let $p \in [0, \infty)$. A mapping $\rho : \mathcal{M}_p \to \mathbb{R}$ satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_1$-cEMA if and only if $\rho = \mathrm{VaR}_\alpha$ for some $\alpha \in (0, 1)$.*

To prove Theorem EC.1, we need some technical lemmas. The following lemma shows that a $\preceq_1$-consistent and lower semicontinuous risk measure can be uniquely extended from $\mathcal{M}_\infty$ to $\mathcal{M}_{\mathrm{bb}}$.

Lemma EC.2. *Let $p \in [0, \infty)$ and $\rho_1, \rho_2 : \mathcal{M}_p \to \mathbb{R} \cup \{\infty\}$ be two $\preceq_1$-consistent and lower semicontinuous risk measures that coincide on $\mathcal{M}_\infty$. Then $\rho_1(F) = \rho_2(F)$ for all $F \in \mathcal{M}_{\mathrm{bb}} \cap \mathcal{M}_p$.*

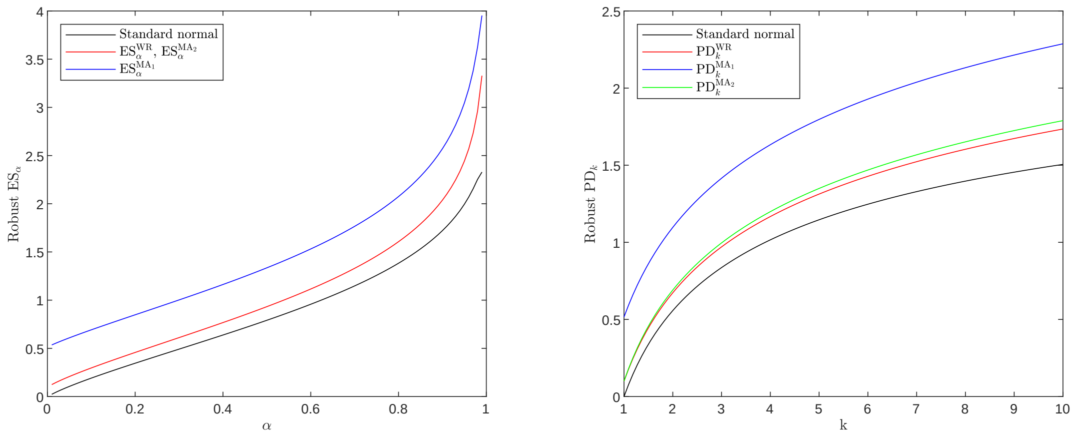
Proof. For $F \in \mathcal{M}_{\mathrm{bb}} \cap \mathcal{M}_p$, without loss of generality assume that the support of F is bounded from below by $-M$. Define $F_n(x) = \sum_{k=0}^{n2^n} F(k/2^n - M)\mathbb{1}_{\{\frac{k-1}{2^n} \leqslant x+M < \frac{k}{2^n}\}} + \mathbb{1}_{\{x+M \geqslant n\}}$, $n \in \mathbb{N}$. We have $F_n \in \mathcal{M}_\infty$, $F_n \preceq_1 F_{n+1}$ for all $n \geqslant 1$, $F_n \xrightarrow{\mathrm{d}} F$ as $n \to \infty$, and

$$\rho_1(F) \geqslant \limsup_{n\to\infty} \rho_1(F_n) = \limsup_{n\to\infty} \rho_2(F_n) \geqslant \liminf_{n\to\infty} \rho_2(F_n) \geqslant \rho_2(F),$$

where the first inequality follows from the $\preceq_1$-consistency of ρ_1, and the last inequality is due to the lower semicontinuity of ρ_2. By symmetry, we have $\rho_1 = \rho_2$. □

Denote by ℓ the Lebesgue measure on $[0, 1]$ and $\mathcal{M}_{1,f}(\ell)$ the space of all finitely additive probability measures on $([0, 1], \mathfrak{B}([0, 1]))$ that are absolutely continuous with respect to ℓ. By Theorem 4.5 of Jia et al. (2020), for any $\preceq_1$-consistent, translation invariant and positively homogeneous

Figure EC.3: The WR and MA approaches under $\mathcal{F}_{2,0.1}(F_0)$ with $F_0 \sim N(0,1)$.



risk measure $\rho : \mathcal{M}_\infty \to \mathbb{R}$, there exists a family $\{\mathcal{M}_\xi : \xi \in \Xi\}$ (Ξ is an index set) of nonempty, weak*-compact and convex subsets of $\mathcal{M}_{1,f}(\ell)$ such that

$$\rho(F) = \min_{\xi \in \Xi} \max_{\mu \in \mathcal{M}_\xi} \int_0^1 \mathrm{VaR}_s(F)\mu(\mathrm{d}s), \quad F \in \mathcal{M}_\infty. \tag{EC.17}$$

Applying this representation, we can establish the following lemma.

Lemma EC.3. *Let $\rho : \mathcal{M}_p \to \mathbb{R}$, $p \in [0,\infty)$, be a mapping satisfying $\preceq_1$-consistency, translation invariance, positive homogeneity and lower semicontinuity. Denote by $\widehat{\rho}$ the restriction of ρ on $\mathcal{M}_\infty$, i.e., $\widehat{\rho}(F) = \rho(F)$ for all $F \in \mathcal{M}_\infty$. Then the following three statements are equivalent.*

(a) $\widehat{\rho} = \mathrm{VaR}_\alpha$ for some $\alpha \in (0,1)$ on $\mathcal{M}_\infty$.

(b) $\widehat{\rho}$ satisfies $\preceq_1$-cEMA.

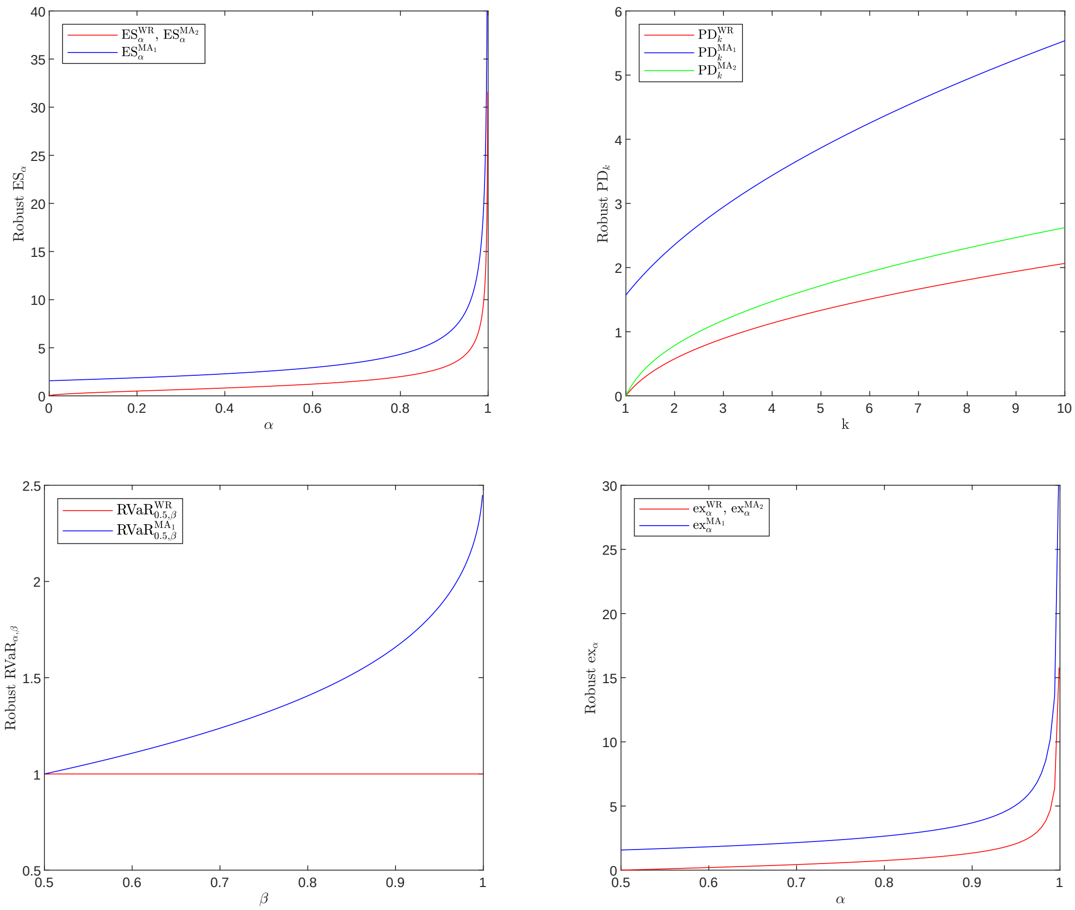
(c) $\widehat{\rho}$ admits a representation (EC.17) which satisfies $\max_{\xi\in\Xi} \min_{\mu\in\mathcal{M}_\xi} \mu([0,s]) = \mathbb{1}_{\{s \geqslant \alpha\}}$ for some $\alpha \in (0,1)$.

Proof. The implication (a)⇒(b) is straightforward to check by Proposition EC.1.

(b)⇒(c): By the discussion before this lemma, there exists a family $\{\mathcal{M}_\xi : \xi \in \Xi\}$ consisting of nonempty, weak*-compact and convex subsets of $\mathcal{M}_{1,f}(\ell)$ such that $\widehat{\rho}$ admits a representation (EC.17). Denote by

$$g_\mu(\beta) = \mu([0,\beta]), \ \beta \in [0,1], \quad g_\xi = \min_{\mu \in \mathcal{M}_\xi} g_\mu \ \text{ and } \ g_\Xi = \max_{\xi \in \Xi} g_\xi.$$

Figure EC.4: The WR and MA approaches under mean-variance uncertainty $\mathcal{F}_{0,1}$.



All these three functions are nonnegative, increasing, and take value one at 1. We complete the proof of (b)⇒(c) by verifying the following three facts.

1. *Right-continuity of* g_Ξ. Note that for any $\beta \in [0,1)$,

$$\begin{aligned}\rho(\beta\delta_0 + (1-\beta)\delta_1) &= \min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi}\int_0^1 \mathrm{VaR}_s(\beta\delta_0+(1-\beta)\delta_1)\mu(\mathrm{d}s)\\ &= \min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi}\int_{(\beta,1]} 1\mu(\mathrm{d}s)\\ &= \min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi}\mu((\beta,1]) = \min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi}(1-g_\mu(\beta)) = 1-g_\Xi(\beta). \qquad \text{(EC.18)}\end{aligned}$$

Fix $\beta \in [0,1)$. Let $\{\beta_n\}_{n\in\mathbb{N}} \subseteq [0,1)$ such that $\beta_n > \beta$ and $\beta_n \downarrow \beta$ as $n\to\infty$. We have $\beta_n\delta_0 + (1-\beta_n)\delta_1 \preceq_1 \beta\delta_0 + (1-\beta)\delta_1$, $n\in\mathbb{N}$, and $\beta_n\delta_0+(1-\beta_n)\delta_1 \xrightarrow{\mathrm{d}} \beta\delta_0+(1-\beta)\delta_1$ as

$n \to \infty$. Therefore, by (EC.18), we have

$$\limsup_{n\to\infty}\{1 - g_\Xi(\beta_n)\} = \limsup_{n\to\infty} \rho(\beta_n\delta_0 + (1-\beta_n)\delta_1) \leqslant \rho(\beta\delta_0 + (1-\beta)\delta_1) = 1 - g_\Xi(\beta),$$

where the inequality comes from $\beta_n\delta_0 + (1-\beta_n)\delta_1 \preceq_1 \beta\delta_0 + (1-\beta)\delta_1$ and $\preceq_1$-consistency of ρ. By lower semicontinuity of ρ, we have

$$\liminf_{n\to\infty}\{1 - g_\Xi(\beta_n)\} = \liminf_{n\to\infty} \rho(\beta_n\delta_0 + (1-\beta_n)\delta_1) \geqslant \rho(\beta\delta_0 + (1-\beta)\delta_1) = 1 - g_\Xi(\beta).$$

Hence, we obtain $g_\Xi(\beta) = \lim_{n\to\infty} g_\Xi(\beta_n)$ which implies the right-continuity of g_Ξ.

2. $g_\Xi(1-) = g_\Xi(1) = 1$. Assume by contradiction that $g_\Xi(1-) = 1 - \delta$ with $\delta \in (0, 1]$. By the definition of g_Ξ, we obtain that for all $\xi \in \Xi$, there exists $\mu \in \mathcal{M}_\xi$ such that $g_\mu(1-) \leqslant g_\Xi(1-) = 1 - \delta$. Take $G \in \mathcal{M}_p \setminus \mathcal{M}_\infty$, $p \in [0, \infty)$ with support on $\mathbb{R}_+$. For any $M \in \mathbb{R}_+$, take $\beta = G(M) \in [0, 1)$ and define $F := \beta\delta_0 + (1-\beta)\delta_M \in \mathcal{M}_\infty$. One can verify that $F \geqslant G$ and thus, $F \preceq_1 G$. Therefore,

$$\rho(G) \geqslant \rho(F) = \widehat{\rho}(F) = \min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi} \int_0^1 \mathrm{VaR}_s(F)\mu(\mathrm{d}s) = M(1 - g_\Xi(\beta)) \geqslant \delta M,$$

where the last inequality follows from $g_\Xi(\beta) \leqslant g_\Xi(1-) = 1 - \delta$ for any $\beta < 1$. Since M is arbitrary, this yields a contradiction to that $\rho : \mathcal{M}_p \to \mathbb{R}$, $p \in [0, \infty)$.

3. $g_\Xi(\beta) = \mathbb{1}_{\{\beta\geqslant\alpha\}}$ *for some* $\alpha \in (0, 1)$. Define

$$\alpha = \inf\{\beta : g_\Xi(\beta) > 0\} \in [0, 1]. \tag{EC.19}$$

We assert $\alpha \in (0, 1)$. To see this, first note that $g_\Xi(1-) = 1$ which implies $\alpha < 1$. We next show $\alpha > 0$ by contradiction. Suppose that $\alpha = 0$, which means $g_\Xi(\beta) > 0$ for each $\beta > 0$. Let $F = \beta\delta_{-a} + (1-\beta)\delta_1$ and $G = \delta_0$ with $\beta \in (0, 1)$ and $a > 0$. We can calculate the MA robust risk value

$$\widehat{\rho}\left(\bigvee_1\{F, G\}\right) = \widehat{\rho}(\beta\delta_0 + (1-\beta)\delta_1) = 1 - g_\Xi(\beta),$$

and the WR robust risk value

$$\begin{aligned}
\sup_{\lambda\in[0,1]} \widehat{\rho}(\lambda F+(1-\lambda)G) &= \sup_{\lambda\in[0,1]} \widehat{\rho}(\lambda\beta\delta_{-a}+(1-\lambda)\delta_0+\lambda(1-\beta)\delta_1)\\
&= \sup_{\lambda\in[0,1]} \min_{\xi\in\Xi} \max_{\mu\in\mathcal{M}_\xi} \{-ag_\mu(\lambda\beta)+1-g_\mu(1-\lambda(1-\beta))\}\\
&= 1-\inf_{\lambda\in[0,1]} \max_{\xi\in\Xi} \min_{\mu\in\mathcal{M}_\xi} \{ag_\mu(\lambda\beta)+g_\mu(1-\lambda(1-\beta))\}\\
&\leqslant 1-\inf_{\lambda\in[0,1]} \max_{\xi\in\Xi}\{a\min_{\mu\in\mathcal{M}_\xi} g_\mu(\lambda\beta)+\min_{\mu\in\mathcal{M}_\xi} g_\mu(1-\lambda(1-\beta))\}\\
&= 1-\inf_{\lambda\in[0,1]} \max_{\xi\in\Xi}\{a\, g_\xi(\lambda\beta)+g_\xi(1-\lambda(1-\beta))\}.
\end{aligned}$$

The property $\preceq_1$-cEMA implies $\widehat{\rho}(\bigvee_1\{F,G\})=\sup_{\lambda\in[0,1]}\widehat{\rho}(\lambda F+(1-\lambda)G)$, and thus,

$$g_\Xi(\beta)\geqslant \inf_{\lambda\in[0,1]} \max_{\xi\in\Xi}\{a\, g_\xi(\lambda\beta)+g_\xi(1-\lambda(1-\beta))\},\quad \beta\in[0,1],\ a>0. \tag{EC.20}$$

Fix $\beta>0$. By (EC.20) and the definition of infimum, for each $a>0$, there exist ε_a and $\lambda_a\in[0,1]$ such that $\lim_{a\to\infty}\varepsilon_a=0$ and

$$g_\Xi(\beta)\geqslant \max_{\xi\in\Xi}\{a\, g_\xi(\lambda_a\beta)+g_\xi(1-\lambda_a(1-\beta))\}-\varepsilon_a. \tag{EC.21}$$

It follows that $g_\Xi(\beta)\geqslant \max_{\xi\in\Xi} a\, g_\xi(\lambda_a\beta)-\varepsilon_a = a\, g_\Xi(\lambda_a\beta)-\varepsilon_a$ which implies $g_\Xi(\lambda_a\beta)\to 0$ as $a\to\infty$. That is, $\lambda_a\to 0$ as $a\to\infty$ since $g(\beta)>0$ for $\beta>0$. Therefore, (EC.21) implies $g_\Xi(\beta)\geqslant \max_{\xi\in\Xi} g_\xi(1-\lambda_a(1-\beta))-\varepsilon_a = g_\Xi(1-\lambda_a(1-\beta))-\varepsilon_a$. Letting $a\to\infty$, we have $g_\Xi(\beta)\geqslant g_\Xi(1-)=1$ for any $\beta>0$ which implies $g_\Xi(s)=\mathbb{1}_{\{s>0\}}$ for $s\in[0,1]$, and this yields a contradiction to that g is right-continuous on $[0,1)$. Hence, we have $\alpha>0$, and thus, $\alpha\in(0,1)$. By the definition of α in (EC.19), we have $g_\Xi(\beta)=0$ for all $\beta\in[0,\alpha)$. Fix $\beta>\alpha$, there exist ε_a and $\lambda_a\in[0,1]$ such that $\lim_{a\to\infty}\varepsilon_a=0$ and (EC.21) holds. Similarly, we have $g_\Xi(\lambda_a\beta)\to 0$ as $a\to\infty$ which implies $\limsup_{a\to\infty}\lambda_a\leqslant \alpha/\beta$. It then follows that $g_\Xi(\beta)\geqslant g_\Xi(1-\lambda_a(1-\beta))-\varepsilon_a$. Letting $a\to\infty$, we have

$$g_\Xi(\beta)\geqslant \limsup_{a\to\infty} g_\Xi(1-\lambda_a(1-\beta))\geqslant g_\Xi\left(\left(1-\frac{\alpha(1-\beta)}{\beta}\right)-\right).$$

Since $\beta<1-\alpha(1-\beta)/\beta$ for $\beta>\alpha$, and the sequence $\{\beta_n\}_{n\in\mathbb{N}}$, where $\beta_0=\beta$ and $\beta_{n+1}=1-\alpha(1-\beta_n)/\beta_n$ for $n\geqslant 0$, converges to 1, and g_Ξ is increasing, we have $g_\Xi(\beta)$ takes constant on $\beta\in(\alpha,1)$, that is, $g_\Xi(\beta)=g_\Xi(1-)=1$. By right-continuity of g_Ξ, we have $g_\Xi(\beta)=\mathbb{1}_{\{\beta\geqslant\alpha\}}$, which completes the proof of (b)⇒(c).

(c)⇒(a): By (EC.18), under the condition of (c), for $F = \beta\delta_0 + (1-\beta)\delta_1$, we have $\rho(F) = 1 - g_\Xi(\beta) = \mathrm{VaR}_\alpha(F)$. By positive homogeneity and translation invariance of ρ, this implies that for any $F = \beta\delta_a + (1-\beta)\delta_b$,

$$\rho(F) = a + (b-a)\min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi}\mu((\beta,1]) = a + (b-a)\mathrm{VaR}_\alpha(\beta\delta_0 + (1-\beta)\delta_1) = \mathrm{VaR}_\alpha(F),$$

where the second equality follows from $\min_{\xi\in\Xi}\max_{\mu\in\mathcal{M}_\xi}\mu((\beta,1]) = 1 - \max_{\xi\in\Xi}\min_{\mu\in\mathcal{M}_\xi}\mu([0,\beta]) = \mathbb{1}_{\{s<\alpha\}}$.

For $F \in \mathcal{M}_\infty$, define $G = \beta\delta_{\mathrm{VaR}_0(F)} + (1-\beta)\delta_{\mathrm{VaR}_\beta(F)}$ for $\beta < \alpha$. One can check $G \preceq_1 F$ and thus, $\rho(F) \geqslant \rho(G) = \mathrm{VaR}_\alpha(G) = \mathrm{VaR}_\beta(F)$ for $\beta < \alpha$. By left-continuity of VaR, we have $\rho(F) \geqslant \lim_{\beta\uparrow\alpha}\mathrm{VaR}_\beta(F) = \mathrm{VaR}_\alpha(F)$. On the other hand, define $H = \alpha\delta_{\mathrm{VaR}_\alpha(F)} + (1-\alpha)\delta_{\mathrm{VaR}_1(F)}$. One can check $F \preceq_1 H$ and thus, $\rho(F) \leqslant \rho(H) = \mathrm{VaR}_\alpha(H) = \mathrm{VaR}_\alpha(F)$. Therefore, we have $\rho(F) = \mathrm{VaR}_\alpha(F)$. □

Proof of Theorem EC.1. Sufficiency follows directly from Proposition EC.1. Below we show necessity. Note that $\preceq_1$-cEMA implies $\preceq_1$-consistency, and hence, by Lemma EC.3, we have $\rho = \mathrm{VaR}_\alpha$ for some $\alpha \in (0,1)$ on $\mathcal{M}_\infty$. It remains to show that $\rho = \mathrm{VaR}_\alpha$ on $\mathcal{M}_p$, $p \in [0,\infty)$. By Lemma EC.2, we have $\rho = \mathrm{VaR}_\alpha$ on $\mathcal{M}_{\mathrm{bb}} \cap \mathcal{M}_p$. Next, we will prove $\rho = \mathrm{VaR}_\alpha$ on $\mathcal{M}_p \setminus \mathcal{M}_{\mathrm{bb}}$. Take $F \in \mathcal{M}_p \setminus \mathcal{M}_{\mathrm{bb}}$ and $\zeta < \mathrm{VaR}_\alpha(F)$. We have

$$\rho\left(\bigvee_1\{F,\delta_\zeta\}\right) = \mathrm{VaR}_\alpha\left(\bigvee_1\{F,\delta_\zeta\}\right) = \mathrm{VaR}_\alpha(F)\vee\zeta = \mathrm{VaR}_\alpha(F), \tag{EC.22}$$

where the first equality is due to $\bigvee_1\{F,\delta_\zeta\} \in \mathcal{M}_{\mathrm{bb}} \cap \mathcal{M}_p$ and the second equality follows from Proposition EC.1. By $F \preceq_1 \bigvee_1\{F,\delta_\zeta\}$ and $\preceq_1$-consistency, we have (EC.22) implies that

$$\rho(F) \leqslant \mathrm{VaR}_\alpha(F) \text{ for all } F \in \mathcal{M}_p \setminus \mathcal{M}_{\mathrm{bb}}. \tag{EC.23}$$

By $\preceq_1$-cEMA of ρ, and together with (EC.22), we have

$$\sup_{0\leqslant\lambda\leqslant1}\rho(\lambda F + (1-\lambda)\delta_\zeta) = \mathrm{VaR}_\alpha(F) \ \text{ for } \zeta < \mathrm{VaR}_\alpha(F). \tag{EC.24}$$

We assert that if F is continuous at $\mathrm{VaR}_\alpha(F)$, then (EC.24) implies

$$\limsup_{\lambda\to1}\rho(\lambda F + (1-\lambda)\delta_\zeta) = \mathrm{VaR}_\alpha(F), \quad \zeta < \mathrm{VaR}_\alpha(F). \tag{EC.25}$$

To see this, fix $\varepsilon > 0$ and $\zeta < \mathrm{VaR}_\alpha(F)$, and let $\lambda \in [0, 1-\varepsilon)$. Since F is continuous at $\mathrm{VaR}_\alpha(F)$, we have $\zeta' := \mathrm{VaR}_{(\alpha-\varepsilon)/(1-\varepsilon)}(F) < \mathrm{VaR}_\alpha(F)$. Hence, $\lambda F + (1-\lambda)\delta_\zeta$ has probability at least $\alpha - \varepsilon + \varepsilon \geqslant \alpha$ for the interval $(-\infty, \zeta' \vee \zeta]$. Therefore, we have $\mathrm{VaR}_\alpha(\lambda F + (1-\lambda)\delta_\zeta) \leqslant \zeta' \vee \zeta < \mathrm{VaR}_\alpha(F)$. Hence,

$$\sup_{0 \leqslant \lambda \leqslant 1-\varepsilon} \rho(\lambda F + (1-\lambda)\delta_\zeta) \leqslant \zeta' \vee \zeta < \mathrm{VaR}_\alpha(F).$$

Therefore, the supremum in (EC.24) is not attained on $[0, 1-\varepsilon)$ for any $\varepsilon > 0$, and we have

$$\limsup_{\lambda \to 1} \rho(\lambda F + (1-\lambda)\delta_\zeta) \vee \rho(F) = \mathrm{VaR}_\alpha(F), \quad \zeta < \mathrm{VaR}_\alpha(F).$$

By lower semicontinuity of ρ, we have

$$\rho(F) \leqslant \liminf_{\lambda \to 1} \rho(\lambda F + (1-\lambda)\delta_\zeta) \leqslant \limsup_{\lambda \to 1} \rho(\lambda F + (1-\lambda)\delta_\zeta).$$

Combining above two equations, the assertion (EC.25) is verified. In the following, we will show that $\rho(G) = \mathrm{VaR}_\alpha(G)$ for $G \in \mathcal{M}_p \setminus \mathcal{M}_{\mathrm{bb}}$ in two cases by applying (EC.23) and (EC.25).

Case 1: *G is continuous at* $\mathrm{VaR}_\alpha(G)$. By (EC.23), $\rho(G) \leqslant \mathrm{VaR}_\alpha(G)$, and we suppose by contradiction that $\rho(G) < \mathrm{VaR}_\alpha(G)$. Since G is continuous at $\mathrm{VaR}_\alpha(G)$, there exist $x^* \in (\rho(G), \mathrm{VaR}_\alpha(G))$ such that G is continuous at x^*, and x^* is an element of the support of G, which implies $\mathrm{VaR}_{G(x^*)}(G) = x^*$. Noting that $x^* < \mathrm{VaR}_\alpha(G)$, we have $\lambda^* := G(x^*)/\alpha \in (0,1)$. Define a cdf as

$$H(x) = \begin{cases} \frac{1}{\lambda^*} G(x) \wedge 1, & x < \mathrm{VaR}_\alpha(G) \\ 1, & x \geqslant \mathrm{VaR}_\alpha(G). \end{cases}$$

One can check that H is continuous at x^* and $\mathrm{VaR}_\alpha(H) = x^*$. By (EC.25), we obtain for $\zeta < x^*$,

$$\limsup_{\lambda \to 1} \rho(\lambda H + (1-\lambda)\delta_\zeta) = \mathrm{VaR}_\alpha(H) = x^*. \tag{EC.26}$$

For $\zeta < x^*$ and $\lambda \in [\lambda^*, 1]$, we have $\lambda H + (1-\lambda)\delta_\zeta \geqslant G$ pointwise, which implies $\lambda H + (1-\lambda)\delta_\zeta \preceq_1 G$. It follows from $\preceq_1$-consistency that $\rho(G) \geqslant \rho(\lambda H + (1-\lambda)\delta_\zeta)$ for all $\zeta < x^*$ and $\lambda \in [\lambda^*, 1]$. Hence, by (EC.26), we obtain $\rho(G) \geqslant x^*$, and this yields a contradiction.

Case 2: *G has a jump at* $\mathrm{VaR}_\alpha(G)$. In this case, we can construct a sequence $\{G_n\}_{n\in\mathbb{N}}$ such that $G_n \preceq_1 F$ for all $n \in \mathbb{N}$, G_n is continuous at point $\mathrm{VaR}_\alpha(G_n)$ and $\mathrm{VaR}_\alpha(G_n) \to \mathrm{VaR}_\alpha(G)$. By Case 1 and $\preceq_1$-consistency of ρ, we have $\rho(G) \geqslant \rho(G_n) \to \mathrm{VaR}_\alpha(G)$. Note that the converse direction holds by (EC.23). Hence, we obtain $\rho(G) = \mathrm{VaR}_\alpha(G)$ for all $F \in \mathcal{M}_p \setminus \mathcal{M}_{\mathrm{bb}}$ such that G has a jump at point ζ.

In summary, we compete the proof of this theorem. $\square$

G.2 A generalization of Theorem 6 (b) and related results

The following theorem is a generalized result of Theorem 6 (b) for the space $\mathcal{M}_p$, $p \in [1, \infty)$.

Theorem EC.2. *A mapping $\rho : \mathcal{M}_p \to \mathbb{R}$, $p \in [1, \infty)$, satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_2$-cEMA if and only if $\rho = \mathrm{ES}_\alpha$ for some $\alpha \in (0, 1)$.*

By Theorem 1, we know that ES satisfies $\preceq_2$-cEMA. In order to prove the necessity of Theorem EC.2, we need to apply Corollary 5.9 of Jia et al. (2020). Define

$$\mathcal{H} := \{h : [0, 1] \to [0, 1] : \ h \text{ is increasing convex}, \ h(0) = 0, \ h(1) = 1\}.$$

For any $\preceq_2$-consistent, translation invariant and positively homogeneous risk measure $\rho : \mathcal{M}_\infty \to \mathbb{R}$, there exists a family $\{\mathcal{H}_\xi : \xi \in \Xi\}$ of nonempty, compact and convex subsets of $\mathcal{H}$ such that

$$\rho(F) = \min_{\xi \in \Xi} \max_{h \in \mathcal{H}_\xi} \int_0^1 \mathrm{VaR}_s(F) \mathrm{d}h(s), \quad F \in \mathcal{M}_\infty. \tag{EC.27}$$

Here, compactness is defined with respect to the weak topology induced by all continuous functions on $[0, 1]$. As pointed out by Jia et al. (2020), both the min and max can be attained, that is, for each $F \in \mathcal{M}_\infty$, there exists $h \in \mathcal{H}_\xi$ for some $\xi \in \Xi$ such that $\rho(F) = \int_0^1 \mathrm{VaR}_s(F)\mathrm{d}h(s)$. Therefore, for any $\beta \in [0, 1]$, by $\rho(\beta\delta_0 + (1 - \beta)\delta_1) = \min_{\xi \in \Xi} \max_{h \in \mathcal{H}_\xi} \int_\beta^1 1 \mathrm{d}h(s) = 1 - \max_{\xi \in \Xi} \min_{h \in \mathcal{H}_\xi} h(\beta)$, we can define

$$h_\xi(\beta) = \min_{h \in \mathcal{H}_\xi} h(\beta) \ \text{ and } \ h_\Xi(\beta) = \max_{\xi \in \Xi} h_\xi(\beta), \quad \beta \in [0, 1]. \tag{EC.28}$$

Applying this representation, we can establish the following lemma.

Lemma EC.4. *Let $\rho : \mathcal{M}_p \to \mathbb{R}$, $p \in [1, \infty)$, be a mapping satisfying $\preceq_2$-consistency, translation invariance and positive homogeneity. Denote by $\widehat{\rho}$ the restriction of ρ on $\mathcal{M}_\infty$. Then the following three statements are equivalent.*

(a) $\widehat{\rho} = \mathrm{ES}_\alpha$ *for some* $\alpha \in [0, 1)$ *on* $\mathcal{M}_\infty$.

(b) $\widehat{\rho}$ *satisfies* $\preceq_2$*-cEMA.*

(c) $\widehat{\rho}$ *admits a representation* (EC.27) *which satisfies* $h_\Xi(\beta) = (\beta - \alpha)_+/(1 - \alpha)$ *for some* $\alpha \in [0, 1)$ *where* h_Ξ *is defined by* (EC.28).

Proof. The implication (a)⇒(b) follows immediately from Theorem 1 (b).

(b)⇒(c): By the discussion before this lemma, there exists a family $\{\mathcal{H}_\xi : \xi \in \Xi\}$ of nonempty, compact and convex subsets of $\mathcal{H}$ such that $\widehat{\rho}$ admits a representation (EC.27). Define h_ξ and h_Ξ by (EC.28). One can check that both h_ξ and h_Ξ are increasing and satisfy $h_\xi(0) = h_\Xi(0) = 0$ and $h_\xi(1) = h_\Xi(1) = 1$. In the following, we show (c) by verifying the following facts.

1. $h_\Xi(1-) = h_\Xi(1) = 1$. This can be showed by similar arguments as in (b)⇒(c) of Lemma EC.3.
2. $h_\Xi(\beta) = (\beta-\alpha)_+/(1-\alpha)$ *for some* $\alpha \in [0,1)$. Let $F = \beta\delta_{-a} + (1-\beta)\delta_1$ and $G = \delta_0$ with $\beta \in (0,1)$ and $a > (1-\beta)/\beta$, and calculate the MA robust risk value

$$\widehat{\rho}\left(\bigvee_2 \{F, G\}\right) = \widehat{\rho}\left(\beta\delta_{-\frac{1-\beta}{\beta}} + (1-\beta)\delta_1\right) = 1 - \frac{h_\Xi(\beta)}{\beta},$$

and the WR robust risk value

$$\begin{aligned}
\sup_{\lambda\in[0,1]} \widehat{\rho}\,(\lambda F + (1-\lambda)G) &= \sup_{\lambda\in[0,1]} \widehat{\rho}(\lambda\beta\delta_{-a} + (1-\lambda)\delta_0 + \lambda(1-\beta)\delta_1) \\
&= \sup_{\lambda\in[0,1]} \min_{\xi\in\Xi} \max_{h\in\mathcal{H}_\xi} \{-a\, h(\lambda\beta) + 1 - h(1-\lambda(1-\beta))\} \\
&= 1 - \inf_{\lambda\in[0,1]} \max_{\xi\in\Xi} \min_{h\in\mathcal{H}_\xi} \{a\, h(\lambda\beta) + h(1-\lambda(1-\beta))\} \\
&\leqslant 1 - \inf_{\lambda\in[0,1]} \max_{\xi\in\Xi} \{a \min_{h\in\mathcal{H}_\xi} h(\lambda\beta) + \min_{h\in\mathcal{H}_\xi} h(1-\lambda(1-\beta))\} \\
&= 1 - \inf_{\lambda\in[0,1]} \max_{\xi\in\Xi} \{a\, h_\xi(\lambda\beta) + h_\xi(1-\lambda(1-\beta))\}.
\end{aligned}$$

The property $\preceq_2$-cEMA implies $\sup_{\lambda\in[0,1]} \widehat{\rho}\,(\lambda F + (1-\lambda)G) = \widehat{\rho}\,(\bigvee_2\{F,G\})$, and thus,

$$\frac{h_\Xi(\beta)}{\beta} \geqslant \inf_{\lambda\in[0,1]} \max_{\xi\in\Xi} \{a\, h_\xi(\lambda\beta) + h_\xi(1-\lambda(1-\beta))\}, \quad \beta\in[0,1],\ a > \frac{1-\beta}{\beta}. \tag{EC.29}$$

Define

$$\alpha = \inf\{\beta : h_\Xi(\beta) > 0\} \in [0,1), \tag{EC.30}$$

where the fact $\alpha < 1$ comes from $h_\Xi(1-) = 1$. Fix $\beta > \alpha$. By (EC.29) and the definition of infimum, for each $a > (1-\beta)/\beta$, there exist $\lambda_a \in [0,1]$ and ε_a such that $\lim_{a\to\infty} \varepsilon_a = 0$ and

$$\frac{h_\Xi(\beta)}{\beta} \geqslant \max_{\xi\in\Xi} \{a h_\xi(\lambda_a\beta) + h_\xi(1-\lambda_a(1-\beta))\} - \varepsilon_a. \tag{EC.31}$$

It follows that $h_\Xi(\beta)/\beta \geqslant \max_{\xi\in\Xi} a\, h_\xi(\lambda_a\beta) - \varepsilon_a = a\, h_\Xi(\lambda_a\beta) - \varepsilon_a$. Letting $a\to\infty$, we obtain

$\lim_{a\to\infty} h_\Xi(\lambda_a\beta)=0$. By definition of α, $\limsup_{a\to\infty}\lambda_a \leqslant \alpha/\beta$. Again, by (EC.31), we have

$$\frac{h_\Xi(\beta)}{\beta} \geqslant \max_{\xi\in\Xi}\{h_\xi(1-\lambda_a(1-\beta))\}-\varepsilon_a = h_\Xi(1-\lambda_a(1-\beta))-\varepsilon_a \tag{EC.32}$$

By monotonicity of h_Ξ and $\limsup_{a\to\infty}\lambda_a \leqslant \alpha/\beta$, we get $\limsup_{a\to\infty} h_\Xi(1-\lambda_a(1-\beta))-\varepsilon_a \geqslant h_\Xi((1-\alpha(1-\beta)/\beta)-)$. This combined with (EC.32) implies $h_\Xi(\beta)/\beta \geqslant h_\Xi((1-\alpha(1-\beta)/\beta)-)$. Denote by $h_\Xi^-(x)=\lim_{y\uparrow x} h_\Xi(y)$. We have $h_\Xi^-(\alpha)=0$ and

$$\frac{h_\Xi^-(\beta)-h_\Xi^-(\alpha)}{\beta-\alpha} \geqslant \frac{h_\Xi^-(\alpha+(1-\alpha/\beta))-h_\Xi^-(\alpha)}{1-\alpha/\beta}, \quad \beta>\alpha. \tag{EC.33}$$

Letting $\beta_0=\beta$ and $\beta_{n+1}=1+\alpha-\alpha/\beta_n$ for $n\geqslant 0$, we have $\beta_n=\frac{\alpha(1-\beta)+(\beta-\alpha)\alpha^{-n+1}}{(1-\beta)+(\beta-\alpha)\alpha^{-n+1}} \uparrow 1$ as $n\to\infty$. Combining with (EC.33), we obtain

$$\frac{h_\Xi^-(\beta)-h_\Xi^-(\alpha)}{\beta-\alpha} \geqslant \frac{h_\Xi^-(\beta_n)-h_\Xi^-(\alpha)}{\beta_n-\alpha} \to \frac{h_\Xi^-(1)-h_\Xi^-(\alpha)}{1-\alpha} = \frac{1}{1-\alpha} \quad \text{as} \quad n\to\infty$$

for all $\beta\in(\alpha,1]$. It follows that $h_\Xi(\beta)\geqslant h_\Xi^-(\beta)\geqslant(\beta-\alpha)_+/(1-\alpha)$ for $\beta\in(\alpha,1]$. We next show $h_\Xi(\beta)\leqslant(\beta-\alpha)_+/(1-\alpha)$ for $\beta\in(\alpha,1]$ by contradiction. Suppose that there exists $\beta^*\in(\alpha,1)$ such that $h_\Xi(\beta^*)>(\beta^*-\alpha)_+/(1-\alpha)$. Noting that $h_\Xi(\beta)=\max_{\xi\in\Xi}\min_{h\in\mathcal{H}_\xi} h(\beta)$, there exists $\xi_0\in\Xi$ such that

$$\min_{h\in\mathcal{H}_{\xi_0}} h(\beta^*) > \frac{(\beta^*-\alpha)_+}{1-\alpha}. \tag{EC.34}$$

Meanwhile, by $h_\Xi(\beta)=0$ for $\beta<\alpha$, we have $\min_{h\in\mathcal{H}_\xi} h(\beta)=0$ for any $\beta<\alpha$ and any $\xi\in\Xi$, and thus, $\min_{h\in\mathcal{H}_{\xi_0}} h(\beta)=0$ for $\beta<\alpha$. This implies that there exists $h_0\in\mathcal{H}_{\xi_0}$ such that $h_0(\beta)=0$ for $\beta<\alpha$. By (EC.34), we have $h_0(\beta^*)>(\beta^*-\alpha)_+/(1-\alpha)$, which, combined with $h_0(1)=1$ and $h_0(\beta)=0$ for $\beta<\alpha$, yields a contradiction to that $h_0\in\mathcal{H}$ is convex. Hence, $h_\Xi(\beta)=(\beta-\alpha)_+/(1-\alpha)$, and this completes the proof of (b)⇒(c).

(c)⇒(a): One can check that under the condition of (c), for $F=\beta\delta_0+(1-\beta)\delta_1$, we have $\rho(F)=1-h_\Xi(\beta)=\mathrm{ES}_\alpha(F)$. By positive homogeneity and translation invariance of ρ, for any $F=\beta\delta_a+(1-\beta)\delta_b$, $\rho(F)=\mathrm{ES}_\alpha(F)$. For $F\in\mathcal{M}_\infty$, define $G=\alpha\delta_{\mathrm{VaR}_0(F)}+(1-\alpha)\delta_{\mathrm{ES}_\alpha(F)}$. By computing the π function, one can check $G\preceq_2 F$ and thus, $\rho(F)\geqslant\rho(G)=\mathrm{ES}_\alpha(G)=\mathrm{ES}_\alpha(F)$. On the other hand, define $H=\beta\delta_{\mathrm{VaR}_\alpha(F)}+(1-\beta)\delta_{\mathrm{VaR}_1(F)}$, where $\beta>\alpha$ satisfies $(\beta-\alpha)\mathrm{VaR}_\alpha(F)+(1-\beta)\mathrm{VaR}_1(F)=(1-\alpha)\mathrm{ES}_\alpha(F)$, that is, $\mathrm{ES}_\alpha(H)=\mathrm{ES}_\alpha(F)$. By computing the ES_s, $s\in[0,1]$, one can check $F\preceq_2 H$ and thus, $\rho(F)\leqslant\rho(H)=\mathrm{ES}_\alpha(H)=\mathrm{ES}_\alpha(F)$. We therefore have $\rho(F)=\mathrm{ES}_\alpha(F)$, which completes the proof. □

Proof of Theorem EC.2. Translation invariance, positive homogeneity and lower semicontinuity of ES_α, $\alpha \in (0,1)$, are well-known, and the property $\preceq_2$-cEMA of ES follows from Theorem 1. Conversely, note that $\preceq_2$-cEMA implies $\preceq_2$-consistency, and hence, by Lemma EC.4, we have $\rho = \mathrm{ES}_\alpha$ for some $\alpha \in [0,1)$ on $\mathcal{M}_\infty$. Thus, it remains to show that this representation can be extended to $\mathcal{M}_p$ for $p \in [1,\infty)$. To see this, for $F \in \mathcal{M}_p$, let X be a random variable with cdf F. Since the probability space is nonatomic, there exists a uniform random variable U on $[0,1]$ such that $X = F^{-1}(U)$ $\mathbb{P}$-a.s. (see, e.g., Lemma A.28 of Föllmer and Schied (2016)). Define

$$U_n = \sum_{i=0}^{n-1} \frac{\alpha i}{n} \mathbb{1}_{\left\{\frac{\alpha i}{n} \leqslant U < \frac{\alpha(i+1)}{n}\right\}} + \sum_{i=0}^{n-1} \left(\alpha + \frac{(1-\alpha)i}{n}\right) \mathbb{1}_{\left\{\alpha + \frac{(1-\alpha)i}{n} \leqslant U < \alpha + \frac{(1-\alpha)(i+1)}{n}\right\}}, \quad n \geqslant 1.$$

and denote by $X_n = \mathbb{E}[X|U_n]$. One can obatin $F_{X_n} \in \mathcal{M}_\infty$, and $\rho(F_{X_n}) = \mathrm{ES}_\alpha(F_{X_n}) = \mathrm{ES}_\alpha(F)$. On one hand, since $F_{X_n} \preceq_2 F$ for all $n \geqslant 1$, and note that $\preceq_2$-cEMA implies $\preceq_2$-consistency, we have $\rho(F_{X_n}) \leqslant \rho(F)$. Hence, we have $\mathrm{ES}_\alpha(F) = \limsup_{n\to\infty} \rho(F_{X_n}) \leqslant \rho(F)$. On the other hand, noting that $X_n \xrightarrow{\mathrm{d}} X$, it follows from the lower semicontinuitycof ρ that $\mathrm{ES}_\alpha(F) = \liminf_{n\to\infty} \rho(F_{X_n}) \geqslant \rho(F)$. Hence, we conclude that $\rho(F) = \mathrm{ES}_\alpha(F)$. Since $\mathrm{ES}_0 = \mathbb{E}$ is not lower semicontinuous, we obtain $\rho = \mathrm{ES}_\alpha$ for some $\alpha \in (0,1)$. □

Remark EC.3. The characterization results in Theorems EC.1 and EC.2 are obtained for spaces $\mathcal{M}_p$, $p \in [1,\infty)$, i.e., cdfs with finite pth moment. On the space $\mathcal{M}_\infty$ of compactly supported cdfs, the situation is more delicate. In particular, for $\alpha \in (0,1)$ and $\lambda \in (0,1)$, we find that the mappings $\lambda \mathrm{VaR}_\alpha + (1-\lambda)\mathrm{VaR}_1$ and $\lambda \mathrm{ES}_\alpha + (1-\lambda)\mathrm{VaR}_1$ on $\mathcal{M}_\infty$ satisfy the conditions in (a) and (b) of Theorem 6, respectively. These mappings are not real-valued on $\mathcal{M}_p$ for $p \in [1,\infty)$. A full characterization on $\mathcal{M}_\infty$ seems beyond current techniques and requires future study. This hints at the level of technical sophistication of the theory.

H Supplementary numerical results in Section 8

We present the summary statistics of the return rates of 20 stocks in Section 8 in Table EC.3.

H.1 Portfolio selection under mean-variance uncertainty

We follow the portfolio selection setting discussed in Section 8.2 to assume that only the mean and the covariance matrix are available to the investor. This appendix complements the study in Section 8.2 with Wasserstein uncertainty.

For a given portfolio weight $\mathbf{w}$, the uncertainty set is $\mathcal{F}_{\mathbf{w}^\top \boldsymbol{\mu}, \sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}}$, where $\boldsymbol{\mu}$ is the mean vector and Σ is the covariance matrix of losses from the stocks as reported in or computed from Table

EC.3. By the results in Section 6.3, the optimization problem of the portfolio selection under the MA approach with $\preceq_1$ and $\preceq_2$ are

$$\min_{\mathbf{w}\in\Delta_{20}} : \quad \rho^{\mathrm{MA}_1}\left(\mathcal{F}_{\mathbf{w}^\top\boldsymbol{\mu},\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}}}\right) = \mathbf{w}^\top\boldsymbol{\mu} + \beta_k\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} \quad \text{s.t. } \mathbf{w}^\top\boldsymbol{\mu} \leqslant -r_0/m, \tag{EC.35}$$

and

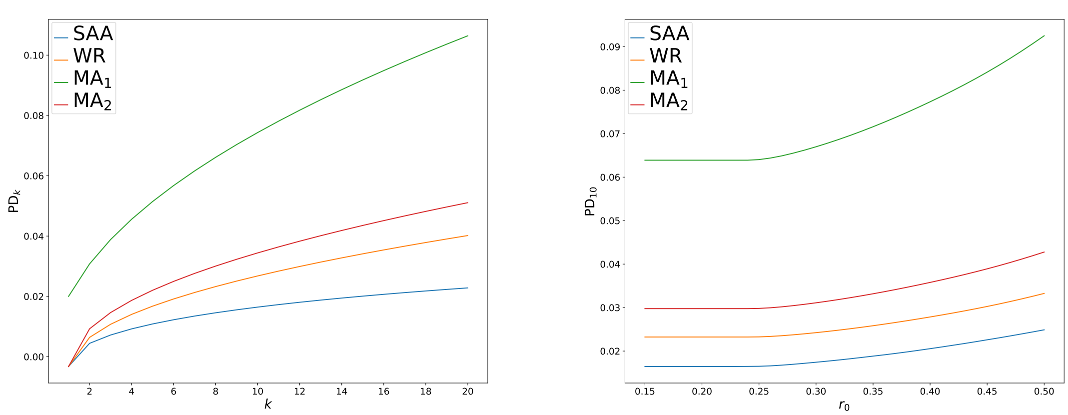
$$\min_{\mathbf{w}\in\Delta_{20}} : \quad \rho^{\mathrm{MA}_2}\left(\mathcal{F}_{\mathbf{w}^\top\boldsymbol{\mu},\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}}}\right) = \mathbf{w}^\top\boldsymbol{\mu} + \gamma_k\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} \quad \text{s.t. } \mathbf{w}^\top\boldsymbol{\mu} \leqslant -r_0/m, \tag{EC.36}$$

respectively, where $\beta_k = (\sqrt{\pi}\Gamma(k+1/2))/\Gamma(k)$ and $\gamma_k = (\sqrt{\pi}(k-1)\Gamma(k+1/2))/((2k-1)\Gamma(k))$, as in Table 1, and r_0 is the expected annualized return and $m = 250$. Using the results of Li (2018), the WR portfolio optimization problem is

$$\min_{\mathbf{w}\in\Delta_{20}} : \quad \rho^{\mathrm{WR}}\left(\mathcal{F}_{\mathbf{w}^\top\boldsymbol{\mu},\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}}}\right) = \mathbf{w}^\top\boldsymbol{\mu} + \eta_k\sqrt{\mathbf{w}^\top\Sigma\mathbf{w}} \quad \text{s.t. } \mathbf{w}^\top\boldsymbol{\mu} \leqslant -r_0/m, \tag{EC.37}$$

where $\eta_k = (k-1)/\sqrt{2k-1}$. Figure EC.5 presents the optimal values of the optimization problem under mean-variance uncertainty with the SAA, WR and MA approaches for different values of k and r_0 using the whole-period data. We can see that the robust value computed by the MA approach with $\preceq_1$ is always the largest one and that of SAA is always the smallest one; this is consistent with our intuition as MA with $\preceq_1$ is the most robust approach among them, and SAA is not conservative.

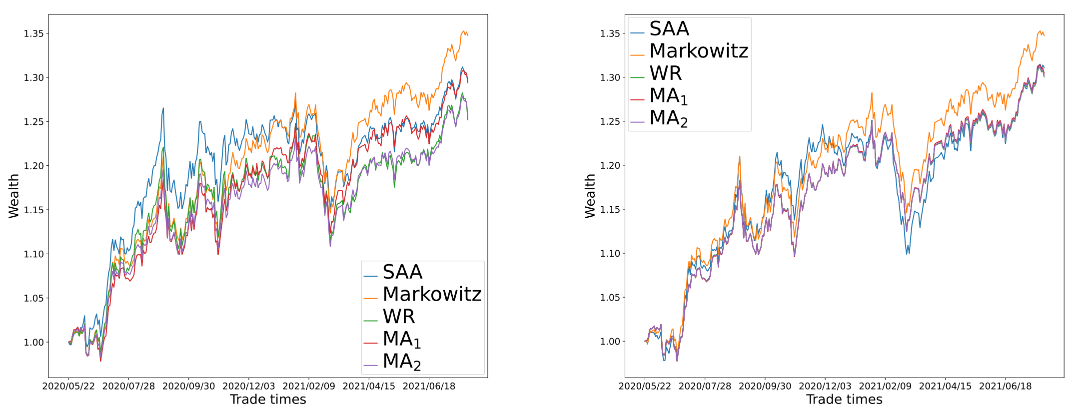
Figure EC.5: The optimized values of PD_k under mean-variance uncertainty. Left: $r_0 = 0.2$ and $k \in [1, 20]$; Right: $k = 10$ and $r_0 \in [0.15, 0.5]$.



Similarly to Section 8.2, we choose 350 trading days for the initial training, and compute the optimal portfolio weights in each day with a rolling window. Figure EC.6 presents the SAA approach, Markowitz model, and WR and MA approaches under Wasserstein uncertainty set with t benchmark distribution and mean-variance uncertainty over the remaining 300 trading days with $r_0 = 0.2$ and $k = 2$ (left) or $k = 20$ (right). In the case $k = 2$, the WR and MA_2 approaches exhibit comparable performance but underperform the other approaches. In the first 150 trading days, the SAA approach outperform the others, and its performance aligns closely with the MA_1 approach thereafter, with the Markowitz model delivering the best performance. In the case $k = 20$, the Markowitz model stands out with superior performance, particularly in the final 100 trading days. Both the MA and WR approaches perform very similarly, trailing slightly behind the SAA approach during the first 150 trading days.

Tables EC.1 and EC.2 present the Sharpe ratios and the nominal transaction cost, respectively, for SAA, Markowitz, WR and MA approaches under mean-variance uncertainty as well as Wasserstein uncertainty, which has been shown in Section 8.2. With $r_0 = 0.2$ and 0.3, the MA and WR approaches based on Wasserstein uncertainty have smaller transaction costs than the other approaches. Additionally, these approaches also exhibit larger Sharpe ratios. The MA_1 approach with mean-variance uncertainty has the smallest transaction cost for $r_0 = 0.1$.

Figure EC.6: Wealth evolution for different portfolio strategies from May 2020 to Aug 2021 with $r_0 = 0.2$. Left: $k = 2$; Right: $k = 20$.



Remark EC.4. The similar performance of MA and WR approaches for large k is not a coincidence. In the setting of mean-variance uncertainty, as k grows, the weights β_k, γ_k and η_k in the optimization problems (EC.35), (EC.36) and (EC.37) also grow. If these weights are large enough, then the

Table EC.1: Annualized return (AR), annualized volatility (AV) and Sharpe ratio (SR) for different portfolio strategies from May 2020 to Aug 2021 with $r_0 = 0.2$

	AR (%)		AV (%)		SR (%)	
Approach	$k=2$	$k=20$	$k=2$	$k=20$	$k=2$	$k=20$
SAA	25.42	26.72	14.82	14.35	170.4	185.0
Markowitz	29.50	29.50	13.54	13.54	216.6	216.6
W-WR	32.75	30.79	14.25	13.30	228.7	230.3
W-MA$_2$	**33.77**	**32.01**	14.48	13.56	**232.0**	**234.9**
MV-WR	22.10	25.60	15.30	12.86	143.4	197.8
MV-MA$_1$	25.26	25.93	**12.92**	**12.81**	194.2	201.1
MV-MA$_2$	22.24	25.71	13.90	12.84	158.8	198.9

Note: W-WR: WR with Wasserstein uncertainty; W-MA$_2$: MA$_2$ with Wasserstein uncertainty; MV-WR: WR with mean-variance uncertainty; MV-MA$_1$: MA$_1$ with mean-variance uncertainty; MV-MA$_2$: MA$_2$ with mean-variance uncertainty.

terms involving $\sqrt{\mathbf{w}^\top \Sigma \mathbf{w}}$ in those problems become dominant in the optimization. For this reason, Problems (EC.35), (EC.36) and (EC.37) are similar to the case that the mixture of mean and standard variance as the objective risk measure with a large weight on the standard variance.

H.2 Wasserstein uncertainty with normal benchmark distribution

We follow the portfolio selection setting discussed in Section 8.2 under Wassserstein uncertainty with a fitted normal benchmark distribution. The considered optimization problems have the same form of (37) and (38) with the unit variance t-distribution replaced by the standard norm

Table EC.2: Nominal transaction cost $\sum_{t=1}^{T} \|\mathbf{w}_{t+1} - \mathbf{w}_t\|_1 / T$ with $\varepsilon = 0.01$ and $T = 299$

	$r_0 = 0.1$		$r_0 = 0.2$		$r_0 = 0.3$	
Approach	$k=2$	$k=20$	$k=2$	$k=20$	$k=2$	$k=20$
SAA	0.0549	0.0071	0.0871	0.0121	0.0969	0.0833
Markowitz	0.0102	0.0102	0.0110	0.0110	0.0746	0.0746
W-WR	0.0127	0.0032	0.0127	**0.0033**	0.0271	0.0382
W-MA$_2$	0.0114	0.0035	**0.0105**	0.0035	**0.0239**	**0.0348**
MV-WR	0.0321	0.0062	0.0766	0.0074	0.0690	0.0714
MV-MA$_1$	**0.0080**	**0.0024**	0.0108	0.0047	0.0714	0.0714
MV-MA$_2$	0.0236	0.0049	0.0496	0.0065	0.0714	0.0714

distribution. Figure EC.7 presents the robust risk values of the optimization problem with the SAA, WR and MA approaches for different values of ε, r_0 and k using the whole-period data. In the left panel, we see that SSA may be the largest if $\varepsilon \leqslant 0.002$, because the empirical cdf of the data may be outside the Wasserstein uncertainty set if ε is too small. As seen from Theorem 4, although the multivariate normal distribution of $\mathbf{X}$ leads to a light-tailed benchmark distribution of $\mathbf{w}^\top\mathbf{X}$, the robust model we use in the MA approach is heavy-tailed. Figure EC.8 reports the wealth process under SAA approach, and WR and MA approaches under Wasserstein uncertainty with a normal benchmark distribution as well as a t-benchmark distribution which has been shown in Section 8.2. All robust approaches perform similarly, and they generally outperform SAA approach, especially after the first 150 trading days.

Figure EC.7: The optimized values of PD_k under Wasserstein uncertainty. Left: $r_0 = 0.2$, $k = 10$ and $\varepsilon \in [0, 0.1]$; Middle: $r_0 = 0.2$, $\varepsilon = 0.01$ and $k \in [1, 20]$; Right: $k = 10$, $\varepsilon = 0.01$ and $r_0 \in [0.15, 0.5]$.

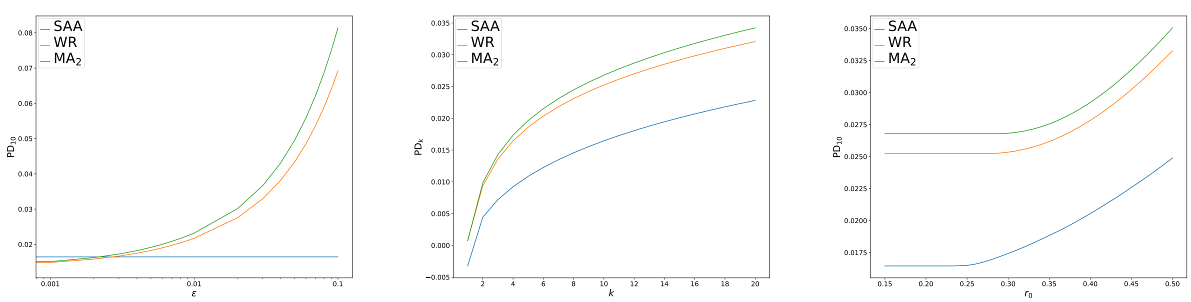
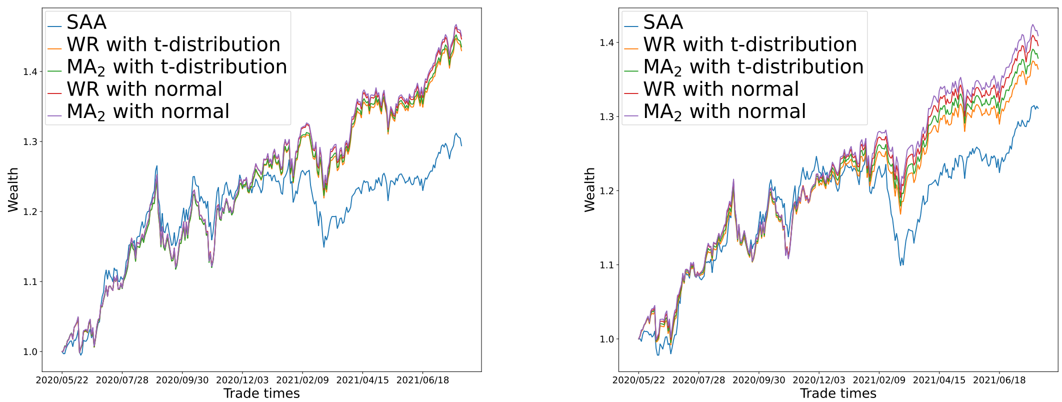


Table EC.3: Summary statistics of daily returns of the 20 stocks, including sample mean ($\times 10^{-3}$), sample standard deviation (SD, $\times 10^{-4}$), and sample correlations

	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	2.3	5.1	1																			
2	1.8	4.0	0.784	1																		
3	1.6	3.8	0.679	0.786	1																	
4	1.4	3.7	0.660	0.714	0.644	1																
5	1.8	5.0	0.707	0.838	0.716	0.704	1															
6	1.3	6.3	0.496	0.546	0.505	0.614	0.612	1														
7	3.2	11.6	0.583	0.600	0.519	0.560	0.574	0.445	1													
8	1.6	3.7	0.637	0.737	0.685	0.443	0.653	0.330	0.473	1												
9	0.7	2.0	0.485	0.563	0.478	0.332	0.429	0.251	0.323	0.568	1											
10	1.3	2.0	0.588	0.647	0.531	0.529	0.600	0.405	0.447	0.513	0.534	1										
11	0.8	2.1	0.438	0.520	0.403	0.378	0.449	0.330	0.334	0.377	0.479	0.680	1									
12	0.9	2.3	0.491	0.581	0.478	0.371	0.494	0.271	0.338	0.538	0.656	0.610	0.610	1								
13	1.4	5.0	0.626	0.715	0.662	0.457	0.628	0.308	0.472	0.922	0.534	0.470	0.349	0.513	1							
14	1.1	4.8	0.500	0.580	0.534	0.357	0.482	0.292	0.379	0.611	0.565	0.483	0.363	0.491	0.567	1						
15	1.0	5.1	0.437	0.493	0.498	0.308	0.402	0.198	0.307	0.649	0.410	0.310	0.274	0.392	0.670	0.415	1					
16	1.3	3.8	0.606	0.681	0.584	0.424	0.598	0.372	0.471	0.654	0.494	0.569	0.469	0.548	0.613	0.593	0.539	1				
17	0.6	6.6	0.584	0.646	0.552	0.461	0.587	0.397	0.434	0.562	0.442	0.514	0.437	0.444	0.548	0.457	0.409	0.560	1			
18	2.1	6.5	0.659	0.742	0.636	0.595	0.742	0.445	0.522	0.653	0.395	0.513	0.374	0.425	0.640	0.428	0.428	0.595	0.540	1		
19	1.6	5.7	0.525	0.544	0.535	0.321	0.440	0.241	0.364	0.649	0.452	0.398	0.321	0.426	0.633	0.556	0.631	0.621	0.521	0.441	1	
20	0.7	3.7	0.479	0.546	0.513	0.354	0.450	0.227	0.361	0.630	0.563	0.455	0.390	0.543	0.623	0.521	0522	0.595	0.555	0.391	0.630	1

Figure EC.8: Wealth evolution for different portfolio strategies from May 2020 to Aug 2021 ($\varepsilon = 0.01$, $r_0 = 0.2$). Left: $k = 2$; Right: $k = 20$.



References

Bäuerle, N. and Müller, A. (2006). Stochastic orders and risk measures: consistency and bounds. *Insurance: Mathematics and Economics*, **38**(1), 132–148.

Berti, P., Pratelli, L. and Rigo, P. (2007). Skorohod representation on a given probability space. *Probability Theory and Related Fields*, **137**(3-4), 277–288.

Cerreia-Vioglio, S., Maccheroni, F., Marinacci, M. and Montrucchio, L. (2011). Risk measures: rationality and diversification. *Mathematical Finance*, **21**(4), 743–774.

Davey, B. A. and Priestley, H. A. (2002). *Introduction to Lattices and Order*, Cambridge University Press, New York.

Föllmer, H. and Schied, A. (2016). *Stochastic Finance. An Introduction in Discrete Time.* (Fourth Edition.) Walter de Gruyter, Berlin.

Huang, M. and Wang, R. (2024). Coherent risk measures and uniform integrability. *arXiv*: 2404.03783.

Jia, G., Xia, J. and Zhao, R. (2020). Monetary risk measures. *arXiv*: 2012.06751.

Kaina, M. and Rüschendorf, L. (2009). On convex risk measures on L^p-spaces. *Mathematical Methods of Operations Research*, **69**(3), 475–495.

Kertz, R. P. and Rösler, U. (2000). Complete lattices of probability measures with applications to martingale theory. *IMS Lecture Notes Monograph Series*, **35**, 153-177.

Rockafellar, R. (1970). *Convex analysis*, Princeton landmarks in mathematics and physics, Princeton University Press.

Liu, F., Mao, T., Wang, R. and Wei, L. (2022). Inf-convolution, optimal allocations, and model uncertainty for tail risk measures. *Mathematics of Operations Research*, **47**(3), 2494–2519.

Shorrocks, A. F. (1983). Ranking income distributions. *Economica*, **50**, 3–17.

Zhu, M. S. and Fukushima, M. (2009). Worst-case conditional value-at-risk with application to robust portfolio management. *Operations Research*, **57**(5), 1155–1168.