Model Aggregation for Risk Evaluation and Robust Optimization

Tiantian Mao\textsuperscript{*} \hspace{1em} Ruodu Wang\textsuperscript{†} \hspace{1em} Qinyu Wu\textsuperscript{‡}

August 16, 2022

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 which is useful for modeling, analysis and simulation, independent of any specific risk measure. The MA approach is easy to implement even if the uncertainty set is non-convex or the risk measure is computationally complicated, and it provides great tractability in distributionally robust optimization. Via an equivalence property between the MA and the WR approaches, new axiomatic characterizations are obtained for a few classes of popular risk measures. In particular, the Expected Shortfall (ES, also known as CVaR) is the unique risk measure satisfying the equivalence property for convex uncertainty sets among a very large class. The MA approach for Wasserstein and mean-variance uncertainty sets admits explicit formulas for the obtained robust models, and 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 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

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

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

\textsuperscript{‡}Department of Statistics and Finance, University of Science and Technology of China, China. E-mail: wu051555@mail.ustc.edu.cn
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 the stressed scenarios and the model aggregation in the FRTB. In the literature of portfolio risk assessment and optimization, the worst-case approach is popular; we refer to 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. 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. This question is not addressed by a worst-case approach of risk evaluation, and it will be addressed by the model aggregation approach that we propose.

We formally describe our main ideas below. Let \(\mathcal{M}\) be the set of all distributions on \(\mathbb{R}\), representing possible risk 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 \textit{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 \(F \in \mathcal{F}\).

\(^1\)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.
(ii) Calibrate a robust (conservative) distributional model $F^*$ from $\mathcal{F}$, and calculate $\rho(F^*)$.

Arguably, each of (i) and (ii) is a reasonable approach to take, but they may yield different 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. Figure 1 contains an illustration of the two methods. The model $F^*$ is robust in two senses: First, it is safer (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 $\rho$ 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 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 what are the implications in optimization and risk management?
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 the follows. After introducing partial orders and risk measures in Section 2, we present a rigorous formulation of the MA and WR approaches for robust risk evaluation in Section 3. Their features and implications in optimization will be discussed in Section 4. We show convenient properties of the MA approach in risk evaluation and optimization, and in particular, the MA approach is more tractable in many settings. This answers Q2, and also Q1 partially.

We establish a few remarkable results in Section 5 that the property of equivalence in model aggregation characterizes VaR, ES, benchmark-loss VaR (Bignozzi et al. (2020)) and benchmark-adjusted ES (Burzoni et al. (2022)) among very general classes of risk measures. The equivalence property is highly desirable and crucially important for optimization, as it identifies for which risk measures the WR approach can be converted to the more tractable MA approach. 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. In particular, the MA approach can conveniently handle multivariate Wasserstein uncertainty in the setting of portfolio selection. Section 7 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 8, which also contains a preliminary discussion on aggregating multivariate risk models, in contrast to the univariate risk models treated throughout the

\(^2\)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., Klíbanoff 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 2 and 3 for other axiomatizations of VaR and ES.
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 collect some notation. Let \((\Omega, \mathcal{B}, \mathbb{P})\) be an atomless probability space, where \(\Omega\) is a set of possible states of nature and \(\mathcal{B}\) is a \(\sigma\)-algebra on \(\Omega\). Let \(L^1\) be the space of all integrable random variables on \((\Omega, \mathcal{B}, \mathbb{P})\), \(M\) a general set of distribution functions, and \(M_1\) the set of the distribution functions of random variables in \(L^1\). We identify distributions with their cumulative distribution functions. The left quantile function of \(F\) is defined by \(F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}\) for \(\alpha \in (0,1]\). We use \(\delta_x\) to represent the point-mass at \(x \in \mathbb{R}\). For a random variable or random vector \(X\), we denote by \(F_X\) the distribution of \(X\).

2.1 Stochastic orders and lattices

As mentioned in the introduction, to properly formulate the MA approach, a partial order \(\preceq\) is needed on \(M\), and \((M, \preceq)\) is called an ordered set. The relevant tool is the lattice theory which we collect in Appendix B, 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 second-order stochastic dominance \(\preceq_2\), defined as, for \(F,G \in M\),

(a) \(F \preceq_1 G\) if \(\int u \, dF \leq \int u \, dG\) for all increasing functions \(u\);

(b) \(F \preceq_2 G\) if \(\int u \, dF \leq \int u \, dG\) for all increasing convex functions \(u\).

Other useful equivalent definitions of \(\preceq_1\) and \(\preceq_2\) are put in Appendix B. To build a robust distributional model, we need to define the supremum of a set \(\mathcal{F}\). For an ordered set \((M, \preceq)\), the supremum of \(\mathcal{F}\), denoted by \(\bigvee \mathcal{F}\), is defined by \(\bigvee \mathcal{F} \in M\) and \(F \preceq_1 \bigvee \mathcal{F} \preceq_1 G\) for all \(F \in \mathcal{F}\) and all \(G \in M\) which dominates every element of \(\mathcal{F}\) (uniqueness is guaranteed by definition). If such \(G\) exists, we say that \(\mathcal{F}\) is **bounded from above**. The supremum does not always exist, but for the two choices of ordered sets \((M_1, \preceq_1)\) and \((M_1, \preceq_2)\) 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.

\(\text{Note that we treat } F \text{ and } G \text{ as loss distributions instead of wealth distributions, and hence a larger element in } \preceq_1 \text{ or } \preceq_2 \text{ means higher risk. Up to a sign change converting losses to gains, } \preceq_1 \text{ and } \preceq_2 \text{ correspond to the classic first- and second-order stochastic dominance in decision theory, respectively.}\)
of distributions with \( \preceq_1 \) and \( \preceq_2 \). Precise definitions and technical details are in Appendix B. In what follows, we use \( \bigvee_1 F \) and \( \bigvee_2 F \) to represent the supremum of the uncertainty set \( F \) on the ordered set \((\mathcal{M}_1, \preceq_1)\) and \((\mathcal{M}_1, \preceq_2)\), respectively, and \( \pi_F \) represents the integrated survival function of \( F \), defined as

\[
\pi_F(x) = \int_{-\infty}^{x} (1 - F(t))dt = \mathbb{E}[\max(0, X - x)], \quad x \in \mathbb{R},
\]

where the random variable \( X \) has distribution \( F \). It is straightforward from (1) that a simple relationship between the integral survival function \( \pi_F \) and the distribution function \( F \) is

\[
F = 1 + (\pi_F)_+,' 
\]

Proposition 1. (i) For a set \( F \subseteq \mathcal{M}_1 \) which is bounded from above with respect to \( \preceq_1 \), we have

\[
\bigvee_1 F = \inf_{F \in F} F
\]

and the left quantile function of \( \bigvee_1 F \) is \( \sup_{F \in F} F^{-1} \).

(ii) For a set \( F \subseteq \mathcal{M}_1 \) which is bounded from above with respect to \( \preceq_2 \), we have

\[
\pi_{\bigvee_2 F} = \sup_{F \in F} \pi_F, \text{ and thus } \bigvee_2 F = 1 + \left( \sup_{F \in F} \pi_F \right)_+.
\]

Remark 1. By Proposition 1, optimization of the supremum with respect to \( \preceq_1 \) leads to the worst-case quantile optimization; see e.g., Ghaoui et al. (2003). Similarly, optimization of the supremum with respect to \( \preceq_2 \) leads to the worst-case first upper partial moment optimization, which also has widely application; see Lo (1987), Natarajan et al. (2010) and Chen et al. (2011).

### 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 \( \mathbb{R} \), and 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.\(^7\) To account for distributional uncertainty in risk management, we work with law-invariant risk measure \( \rho \), that is, identically distributed random variables have the same risk value. For law-invariant risk measures, we can conveniently define them as mappings from a set \( \mathcal{M} \) of distributions to \( \mathbb{R} \), and we will follow this setting; thus, a risk measure is a mapping \( \rho: \mathcal{M} \to \mathbb{R} \) in this paper.

\(^6\)Note that the infimum of upper semicontinuous functions \( F \in F \) is again upper semicontinuous and thus a valid distribution function when \( F \) is bounded from above.

\(^7\)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 distributions.
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 $\text{VaR}_\alpha : \mathcal{M}_1 \rightarrow \mathbb{R}$ defined by
\[
\text{VaR}_\alpha(F) = F^{-1}(\alpha) = \inf\{x \in \mathbb{R} : F(x) \geq \alpha\},
\]
which is the left $\alpha$-quantile of a distribution. The risk measure ES at level $\alpha \in [0, 1)$ is the functional $\text{ES}_\alpha : \mathcal{M}_1 \rightarrow \mathbb{R}$ defined by
\[
\text{ES}_\alpha(F) = \frac{1}{1-\alpha} \int_{\alpha}^{1} \text{VaR}_s(F) ds,
\]
and in particular, $\text{ES}_0(F) = \mathbb{E}[F]$ where $\mathbb{E}$ represents the expectation. We can also define $\text{VaR}_0$, $\text{VaR}_1$ and $\text{ES}_1$ which are not finite-valued on $\mathcal{M}_1$; see Appendix C.

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 $\rho : \mathcal{M} \rightarrow \mathbb{R}$ is $\preceq$-consistent if $\rho(F) \leq \rho(G)$ for all $F, G \in \mathcal{M}$ with $F \preceq G$. For law-invariant mappings, $\preceq_1$-consistency is equivalent to monotonicity. In particular, all monotone law-invariant risk measures, including VaR, are consistent with $\preceq_1$, and ES satisfies both $\preceq_1$ and $\preceq_2$.

3 Introducing the MA approach

We describe the two approaches for robust risk evaluation, the primary objects of this paper. For a risk measure $\rho : \mathcal{M} \rightarrow \mathbb{R}$ 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
\[
\text{WR} : \quad \rho^{\text{WR}}(\mathcal{F}) = \sup_{F \in \mathcal{F}} \rho(F). \tag{3}
\]
The value in (3) is called the WR robust $\rho$ 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 $\bigvee \mathcal{F}$ exists,
\[
\text{MA} : \quad \rho^{\text{MA}}_{\preceq}(\mathcal{F}) = \rho \left( \bigvee \mathcal{F} \right), \tag{4}
\]
and $\rho^{\text{MA}}_{\preceq}(\mathcal{F}) = \infty$ if $\mathcal{F}$ is not bounded from above. The value in (4) is called the $\preceq$-MA robust $\rho$ 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, $\rho^{\text{MA}}_{\preceq}$ is always well-defined. In case that $\bigvee \mathcal{F}$ may not exist, (4) needs to be modified as in Appendix B.

The MA approach is implemented in two steps: First, take the supremum $\bigvee \mathcal{F}$ of the uncer-
tain set $\mathcal{F}$ as the robust distribution, and second, calculate the value of the risk measure of the robust distribution; see Figure 1. The robust distribution $\bigvee \mathcal{F}$ obtained in the first step can be used for any risk measure. If, in addition, the risk measure $\rho$ is $\preceq$-consistent, then the MA approach produces a larger robust risk value than the WR approach, that is, for any $\mathcal{F} \subseteq \mathcal{M}$,

$$\rho_{\text{WR}}(\mathcal{F}) \leq \rho_{\geq 1}(\mathcal{F}),$$

since $\preceq$-consistency implies $\rho(\bigvee \mathcal{F}) \geq \rho(\mathcal{F})$ for all $\mathcal{F} \in \mathcal{F}$. 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 contrast to the MA approach, the WR approach provides a worst-case risk value but not a robust distribution.

If $\rho$ is consistent with more than one partial order, the MA approach with a stronger partial order leads to a higher risk evaluation. For instance, if $\rho$ is both $\preceq_1$-consistent and $\preceq_2$-consistent, then $\rho_{\text{WR}}(\mathcal{F}) \leq \rho_{\geq 2}(\mathcal{F}) \leq \rho_{\geq 1}(\mathcal{F})$, because any $(\mathcal{M}, \preceq_1)$-upper bound on $\mathcal{F}$ is also an $(\mathcal{M}, \preceq_2)$-upper bound on $\mathcal{F}$. In the sequel, we will focus on $\preceq_1$ and $\preceq_2$. For these two stochastic orders, the explicit forms of $\bigvee_1 \mathcal{F}$ or $\bigvee_2 \mathcal{F}$ are simple and obtained in Proposition 1.

4 MA approach for risk measures and robust optimization

4.1 MA approach for convex and non-convex uncertainty sets

Convexity of the uncertainty set is crucial in many optimization problems; see e.g., Zhu and Fukushima (2009) in the case of optimizing ES under uncertainty. For the WR approach, if the uncertainty set is not convex, then the corresponding optimization problem is typically quite difficult to solve. For the MA approach with $\preceq_1$ and $\preceq_2$, the following theorem illustrates a convenient property that the supremum remains if the uncertainty set $\mathcal{F}$ is replaced by its convex hull.

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

Proposition 2 illustrates that to implement the MA approach with $\preceq_1$ or $\preceq_2$, there is no extra difficulty when the uncertainty set is non-convex. This result is straightforward to prove based on Proposition 1 and the fact that for a fixed $x \in \mathbb{R}$, both $F \mapsto F(x)$ and $F \mapsto \pi_F(x)$ are linear in $F$. Despite its simple proof, the result is powerful for many discussions on the MA and the WR approaches in this paper.

As discussed in Remark 1, to obtain $\bigvee_1 \mathcal{F}$ and $\bigvee_2 \mathcal{F}$, one needs to consider the worst-case
quantile and the worst-case first upper partial moment, respectively. Thus, the MA approach with \( \preceq_1 \) or \( \preceq_2 \) is more tractable than the WR approach if the risk measure is itself computationally complicated (or even black-box in some applications),\(^8\) since it requires only a single application of the risk measure to the specific robust model. There are many existing results on the worst-case left quantile and the worst-case first upper partial moment for different uncertainty sets in the literature (see Remark 1), based on which the MA approach can be directly implemented. We will study three concrete settings of uncertainty in Section 6, for which the MA approach admits closed-form formulas for the produced robust model, and the corresponding robust risk value can be easily computed or optimized.

In what follows, we discuss the MA and the WR approaches for the important risk measures VaR and ES, and its properties in optimization. These discussions will shed some light on the advantages of the MA approach, as well as how the two approaches are connected. In particular, the inequality (5) will play an important role in our discussions.

4.2 MA and WR approaches for VaR and ES

We discuss the MA and the WR approaches applied to the risk measures VaR and ES, and this will help us to understand the inequality (5). The case of VaR, coupled with the partial order \( \preceq_1 \), is simple. By Proposition 1, for \( \alpha \in (0, 1) \) and any \( F \) bounded from above, \( \text{VaR}^\alpha_{\text{WR}}(F) = (\text{VaR}_\alpha)^{\text{MA}}_{\preceq_1}(F) \), and thus (5) holds as an equality in this specific setting; this simple 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

\[
\text{ES}^\alpha_{\text{WR}}(F) = \frac{1}{1-\alpha} \sup_{F \in \mathcal{F}} \int_\alpha^1 F^{-1}(s)\,ds \leq \frac{1}{1-\alpha} \int_\alpha^1 \sup_{F \in \mathcal{F}} F^{-1}(s)\,ds = (\text{ES}_\alpha)^{\text{MA}}_{\preceq_1}(F),
\]

for (5) 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 \( F^{-1}(s) \geq G^{-1}(s) \) 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\( _\alpha \) for \( \alpha \in (0, 1) \) in the celebrated work of Rockafellar and Uryasev (2002), that is,

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

---

\( ^8 \)See e.g., Miloslovich et al. (2021) for a discussion on multivariate stress testing, where the risk evaluation procedure may be black-box.
Using (6), we obtain the WR robust ES value, that is,

\[ \text{ES}_\alpha^{WR}(\mathcal{F}) = \sup_{F \in \mathcal{F}} \text{ES}_\alpha(F) = \sup_{F \in \mathcal{F}} \min_{x \in \mathbb{R}} \left\{ x + \frac{1}{1 - \alpha} \pi_F(x) \right\}. \] (7)

On the other hand, the \( \preceq_2 \)-MA robust ES value can also be calculated using (6) and (2) in Proposition 1, that is,

\[ (\text{ES}_\alpha^{MA})_{\preceq_2}(\mathcal{F}) = \text{ES}_\alpha\left( \bigvee_{\preceq_2} \mathcal{F} \right) = \min_{x \in \mathbb{R}} \sup_{F \in \mathcal{F}} \left\{ x + \frac{1}{1 - \alpha} \pi_F(x) \right\}. \] (8)

The formulas (7) and (8) 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

\[ \text{ES}_\alpha^{WR}(\mathcal{F}) \leq (\text{ES}_\alpha^{MA})_{\preceq_2}(\mathcal{F}), \] and equality holds if a minimax theorem holds. \( \tag{9} \)

Therefore, although (5) is generally not an equality, it may be an equality for \( \text{ES}_\alpha \) 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 5.1 for a definition) or a compact convex set of discrete distributions, then (9) becomes an equality. In the following theorem, we establish a more general sufficient condition to make (9) an equality, where \( \text{ES}_0 = \mathbb{E} \) and \( \text{ES}_\alpha \) for \( \alpha \in (0, 1) \) are treated separately. We also collect the corresponding result for VaR\(_\alpha\) discussed above.

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

(a) If \( \sup_{F \in \mathcal{F}} \int_{\mathbb{R}} (x - y)_{+} \, dF(y) \to 0 \) as \( x \to -\infty \), then \( \mathbb{E}^{WR}[\mathcal{F}] = \mathbb{E}^{MA}_{\preceq_2}[\mathcal{F}] \).

(b) For \( \alpha \in (0, 1) \), if \( \mathcal{F} \) is convex and bounded from above with respect to \( \preceq_2 \), then \( \text{ES}_\alpha^{WR}(\mathcal{F}) = (\text{ES}_\alpha)^{MA}_{\preceq_2}(\mathcal{F}) \).

(c) For \( \alpha \in (0, 1) \), if \( \mathcal{F} \) is bounded from above with respect to \( \preceq_1 \), then \( \text{VaR}_\alpha^{WR}(\mathcal{F}) = (\text{VaR}_\alpha)^{MA}_{\preceq_1}(\mathcal{F}) \).

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, typically computationally easier. 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 distributions, respectively. Without convexity of \( \mathcal{F} \), for \( \alpha \in (0, 1) \), \( \text{ES}_\alpha^{WR}(\mathcal{F}) = (\text{ES}_\alpha)^{MA}_{\preceq_2}(\mathcal{F}) \) 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 $F_2 = (1-\varepsilon)\delta_{-1/(1-\varepsilon)} + \varepsilon\delta_{1/\varepsilon}$. By computing $\max\{\pi_{F_1}, \pi_{F_2}\}$, we get $\sqrt{2} \{F_1, F_2\} = (1-\varepsilon)\delta_{-1/(1-\varepsilon)} + \varepsilon\delta_{1/\varepsilon}$ and
\[
\mathbb{E}^\alpha \left( \sqrt{2} \{F_1, F_2\} \right) = \frac{1}{2} \left( \frac{1}{\varepsilon} - \frac{1}{1-\varepsilon} \right) > \frac{1}{2} \left( \frac{1}{\varepsilon} - \frac{2-\varepsilon}{1-\varepsilon} \right) = \max \{ \mathbb{E}^\alpha(F_1), \mathbb{E}^\alpha(F_2) \}.
\]
Hence, $\mathbb{E}^\alpha_{\text{WR}}(\{F_1, F_2\}) < (\mathbb{E}^\alpha_{\text{MA}}(\{F_1, F_2\})$.

The conditions on $\mathcal{F}$ in (a) and (b) of Theorem 1 do not imply each other. The following example shows that $\mathbb{E}^\text{WR}[\mathcal{F}] = \mathbb{E}^\text{MA}[\mathcal{F}]$ 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 $\sqrt{2} \mathcal{F} = \sqrt{2} \{F_n\}_{n \in \mathbb{N}} = \delta_0$. Note that $\mathbb{E}[F] = -1$ for any $F \in \mathcal{F}$. Hence, $\mathbb{E}[\sqrt{2} \mathcal{F}] = 0 > -1 = \sup_{F \in \mathcal{F}} \mathbb{E}[F]$, that is, $\mathbb{E}^\text{WR}(\mathcal{F}) < \mathbb{E}^\text{MA}(\mathcal{F})$.

4.3 MA approach in robust optimization

We consider general robust optimization problems where uncertainty is addressed by the WR and MA approaches. Let $A$ be a set of possible actions and $\mathcal{G}$ be a set of distributions on $\mathbb{R}^d$, where $d \geq 1$. Let $\mathcal{F}_{a,f}$ be the set of univariate distributions of $f(a, X_G)$ where $f : A \times \mathbb{R}^d \to \mathbb{R}$ is a loss function and the random vector $X_G$ has distribution $G \in \mathcal{G}$. For instance, by choosing $A \subseteq \mathbb{R}^d$ and $f(a,x) = a^\top x$, one arrives at the setting of robust portfolio selection, where $a$ represents the vector of portfolio weights and $x$ represents the vector of losses from individual assets.

For a partial order $\preceq$, which is either $\preceq_1$ or $\preceq_2$, we consider the following two optimization problems
\[
\min_{a \in A} \rho^\text{WR}(\mathcal{F}_{a,f}) \quad \text{and} \quad \min_{a \in A} \rho^\text{MA}_{\preceq}(\mathcal{F}_{a,f}). \tag{10}
\]

For (10), we need to solve (i) $\rho^\text{WR}(\mathcal{F}_{a,f})$ and (ii) $\rho^\text{MA}_{\preceq}(\mathcal{F}_{a,f})$ for each fixed $a \in A$. The problem (i) can be quite difficult to solve if the risk measure $\rho$ or the function $f$ is computationally complicated, and results in the literature usually only address a few special cases of $\rho$. The problem (ii) for $\preceq_2$ is more tractable because, as shown in Remark 1, we can equivalently solve a stochastic program,
\[
\max_{G \in \mathcal{G}} : \mathbb{E}[(f(a, X_G) - x)_+], \tag{11}
\]
in which the objective function is linear in $G$. Note that (ii) only requires to evaluate $\rho$ on a robust distribution in (11) for each $a$ whereas (i) requires to repeatedly evaluate $\rho$ on $f(a, X_G)$ for each
distribution \( G \in \mathcal{G} \) and each \( a \); this makes (i) infeasible for many relevant choices of \( \rho \) such as distortion risk measures or rank-dependent expected utilities (Quiggin (1993)).

The problem (11) can be numerically solved for different uncertainty sets; see e.g., Bayraksan and Love (2015) and Esfahani and Kuhn (2018). The portfolio optimization problem in some settings will be investigated in Section 6, where we derive explicit forms of \( \rho^{\text{MA}}_{\leq 1}(F_{a,f}) \) and \( \rho^{\text{MA}}_{\leq 2}(F_{a,f}) \) for general risk measures, by noting that most risk measures and decision criteria are \( \leq 1 \)-consistent, and many are further \( \leq 2 \)-consistent. As the WR optimization problem is difficult to solve, we obtain an upper bound on its robust risk value by solving the more tractable MA optimization problem.

In the special and important case of \( \rho = \text{ES} \), the two optimization problems in (10) are

\[
\min_{a \in A} \text{ES}^{\text{WR}}_\alpha (F_{a,f}) = \min_{a \in A} \sup_{G \in \mathcal{G}} \min_{x \in \mathbb{R}} \left\{ x + \frac{1}{1 - \alpha} \mathbb{E}[f(a, X_G) - x]^+ \right\}, \quad (12)
\]

and

\[
\min_{a \in A} (\text{ES}_\alpha)_{\leq 2}^{\text{MA}} (F_{a,f}) = \min_{a \in A} \sup_{a, x \in \mathbb{R}} \sup_{G \in \mathcal{G}} \left\{ x + \frac{1}{1 - \alpha} \mathbb{E}[f(a, X_G) - x]^+ \right\}. \quad (13)
\]

The value of problem (12) is difficult to compute, while problem (13) is much more tractable as its inner supremum problem is a stochastic program which is widely investigated in the literature. As discussed in Section 4.2, under some conditions, (12) and (13) have equivalent values, but generally they are not identical. Many results in the literature rely on converting between (12) and (13); see e.g., Zhu and Fukushima (2009), Natarajan et al. (2010) and Chen et al. (2011).

If the equality in (5) holds, then one can freely convert between the MA and the WR approaches in robust risk evaluation and optimization. A natural question is which risk measures and uncertainty sets guarantees this equality for \( \leq 1 \) or \( \leq 2 \). This property will be called equivalence in model aggregation (EMA). The general message, which will be clear from our main characterization results in Section 5, is that only a few classes of risk measures may satisfy this property.

5 Characterization of risk measures by equivalence in MA

In this section, we characterize risk measures which satisfy the property of EMA,

\[
\rho \left( \bigvee F \right) = \sup_{F \in F} \rho(F),
\]

for all \( F \in \mathcal{S} \), where \( \mathcal{S} \) is a collection of subsets of \( \mathcal{M} \). We consider two settings where \( \mathcal{S} \) is either the collection of all convex polytopes or the collections of all subsets in \( \mathcal{M} \). The property (14) is
desirable and relevant for optimization, since it allows one to convert optimization problems with the WR formulation to those with the MA formulation, the latter being more tractable as discussed in Section 4.3. Through (14), we obtain new characterization results for the classes of VaR, ES, benchmark-loss VaR in Bignozzi et al. (2020) and benchmark-adjusted ES in Burzoni et al. (2022).

We first collect some standard properties of a risk measure $\rho$, where all distributions involved are assumed to be in the domain of $\rho$.

**Translation invariance:** $\rho(F_{X+c}) = \rho(F_X) + c$ for any $c \in \mathbb{R}$ and random variable $X$.

**Positive homogeneity:** $\rho(F_{\lambda X}) = \lambda \rho(F_X)$ for any $\lambda > 0$ and random variable $X$.

**Convexity:** $\rho(F_{\lambda X + (1-\lambda)Y}) \leq \lambda \rho(F_X) + (1 - \lambda) \rho(F_Y)$ for any $\lambda \in [0, 1]$ and random variables $X$ and $Y$.

**Lower semicontinuity:** $\liminf_{n \to \infty} \rho(F_n) \geq \rho(F)$ if $F_n \overset{d}{\to} F$ as $n \to \infty$, where $\overset{d}{\to}$ denotes weak convergence.\(^\text{9}\)

A risk measure is coherent, as defined by Artzner et al. (1999), if it satisfies monotonicity, translation invariance, positive homogeneity, and convexity.

It is well-known that VaR$^\alpha$ and ES$^\alpha$, $\alpha \in (0, 1)$ satisfy translation invariance, positive homogeneity, lower semicontinuity, and ES$^\alpha$ 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 distribution function is modeled using a truthful approximation, then the approximated risk model should not underreport the capital requirement as the approximation error reduces to zero. A law-invariant coherent risk measure on $\mathcal{M}_1$, including ES, is automatically consistent with both $\preceq_1$ and $\preceq_2$; see e.g., Rüschendorf (2013).

### 5.1 EMA for convex polytopes

We first consider equivalence in model aggregation for convex polytopes (cEMA), that is, (14) holds for all convex polytopes $\mathcal{F} \subseteq \mathcal{M}$. Recall that $\mathcal{F}$ is a convex polytope if it is a convex set generated by finitely many extreme points; that is, there exist finitely many $F_1, \ldots, F_n$ such that

$$\mathcal{F} = \text{conv}(F_1, \ldots, F_n) = \left\{ \sum_{i=1}^n \lambda_i F_i : (\lambda_1, \ldots, \lambda_n) \in \Delta_n \right\},$$

where $\Delta_n = \{(\lambda_1, \ldots, \lambda_n) \in [0, 1]^n : \lambda_1 + \cdots + \lambda_n = 1\}$ is the standard simplex in $\mathbb{R}^n$.

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

\(^\text{9}\)Weak convergence corresponds to convergence in distribution for random variables. 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)).
All results in this section remain valid if convex polytopes in the above definition are replaced by convex sets bounded from above, and such an EMA property is stronger than cEMA.  

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

$$\rho\left(\bigvee_i\{F_1,\ldots,F_n\}\right) = \sup \left\{ \rho\left(\sum_{i=1}^n \lambda_i F_i\right) : (\lambda_1,\ldots,\lambda_n) \in \Delta_n \right\}, \quad i = 1, 2, \tag{15}$$

for all $F_1,\ldots,F_n \in \mathcal{M}$. By (15), $\preceq_i$-cEMA is stronger than $\preceq_i$-consistency since for any $F \preceq_i G$,

$$\rho(G) = \rho\left(\bigvee_i\{F,G\}\right) = \sup_{\lambda \in [0,1]} \rho(\lambda F + (1 - \lambda)G) \geq \rho(F).$$

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, which are remarkable: 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 2.** A mapping $\rho : \mathcal{M}_1 \to \mathbb{R}$ satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_1$-cEMA if and only if $\rho = \text{VaR}_\alpha$ for some $\alpha \in (0,1)$.

*Remark 2.* 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 2, 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.

A characterization of ES via $\preceq_2$-cEMA is obtained in a similar form to Theorem 2.

**Theorem 3.** A mapping $\rho : \mathcal{M}_1 \to \mathbb{R}$ satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_2$-cEMA if and only if $\rho = \text{ES}_\alpha$ for some $\alpha \in (0,1)$.

The special case of $\text{ES}_0 = \mathbb{E}$ is excluded from Theorem 3, as it satisfies $\preceq_2$-cEMA (by Theorem 1) but not lower semicontinuity.

---

10Recall that characterization results are generally stronger if imposed properties are weaker, so we aim for a weaker formulation of the properties.
Remark 3. 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. (2021), 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 3 is quite different from results in the literature and can be applied to robust modeling outside of a financial or statistical context.

Theorems 2 and 3 state that \( \preceq_1 -cEMA \) and \( \preceq_2 -cEMA \) can identify VaR and ES, respectively. In contrast to VaR which satisfies (14) for any \( F \) bounded from above (Theorem 1), ES fails to satisfy (14) for non-convex set \( F \) (Example 1), and hence we need to search for alternative risk measures which yield EMA with \( \preceq_2 \) for arbitrary uncertainty sets, and this will be treated in Section 5.2.

5.2 EMA for arbitrary uncertainty sets

We now move away from the convex polytopes in Section 5.1 and formulate EMA for an arbitrary uncertainty set \( F \subseteq M \).

\( \preceq -EMA \): Let \( (M, \preceq) \) be an ordered set. A mapping \( \rho : M \to \mathbb{R} \) satisfies \( \preceq -EMA \) if \( \rho(\bigvee F) = \sup_{F \in F} \rho(F) \) holds for any \( F \subseteq M \) bounded from above.

The property \( \preceq -EMA \) is stronger than \( \preceq -consistency \) as \( \rho(G) = \rho(\bigvee \{F, G\}) = \max\{\rho(F), \rho(G)\} \) for any \( F \preceq G \). By definition, \( \preceq -EMA \) is also stronger than \( \preceq -cEMA \). With this stronger property, we will relax positive homogeneity used in Section 5.1 to allow for larger classes of risk measures. First, \( \preceq_1 -EMA \) characterizes the class of benchmark-loss VaR in Bignozzi et al. (2020).

Theorem 4. A mapping \( \rho : M_1 \to \mathbb{R} \) satisfies \( \rho(\delta_0) = 0 \), translation invariance, lower semicontinuity and \( \preceq_1 -EMA \) if and only if it is a benchmark-loss VaR, that is,

\[
\rho(F) = \sup_{\alpha \in (0,1)} \{\text{VaR}_\alpha(F) - h(\alpha)\},
\]

for some increasing \( h : (0,1) \to [0, \infty] \) with \( h(0+) = 0 \).

In Theorem 4, \( \rho \) is assumed to be real-valued. Indeed, (16) for an arbitrary \( h \) does not always define a real-valued mapping. A sufficient condition for \( \rho \) in (16) to be real-valued is that \( h(\alpha) = \infty \) for some \( \alpha < 1 \). A technical remark on the conditions of \( h \) is put in Appendix H. Corollary 1 below
immediately follows from Theorem 2 or 4 since $\preceq_1$-EMA is stronger than $\preceq_1$-cEMA and a positively homogeneous benchmark-loss VaR is a VaR (Bignozzi et al. (2020, Proposition 4.6)).

**Corollary 1.** A mapping $\rho : M_1 \to \mathbb{R}$ satisfies translation invariance, positive homogeneity, lower semicontinuity and $\preceq_1$-EMA if and only if $\rho = \text{VaR}_\alpha$ for some $\alpha \in (0, 1)$.

Next, we identify the class of risk measures characterized by $\preceq_2$-EMA on $M_1$. It turns out that these risk measures belong to the class of benchmark-adjusted ES recently introduced by Burzoni et al. (2022). Notably, the obtained class does not include ES, as ES generally fails to satisfy $\preceq_2$-EMA.

**Theorem 5.** A mapping $\rho : M_1 \to \mathbb{R}$ satisfies $\rho(\delta_0) = 0$, translation invariance and $\preceq_2$-EMA if and only if it is a benchmark-adjusted ES, that is,

$$
\rho(F) = \sup_{\alpha \in [0, 1]} \{\text{ES}_\alpha(F) - g(\alpha)\},
$$

for some increasing $g : [0, 1) \to [0, \infty]$ with $g(0+) = 0$ such that $h : \alpha \mapsto (1 - \alpha)g(\alpha)$ is concave on $[0, 1)$ with $h(1-) > 0$.

**Remark 4.** The condition $h(1-) > 0$ is a necessary and sufficient condition for $\rho$ defined by (17) to be real-valued on $M_1$. Burzoni et al. (2022) defined the benchmark-adjusted ES via (17) by choosing $g : \alpha \mapsto \text{ES}_\alpha(G)$ for some $G \in M_1$, that is, $\rho(F) = \sup_{\alpha \in [0, 1]} \{\text{ES}_\alpha(F) - \text{ES}_\alpha(G)\}$. Although sharing the same formula (17), the requirement on $g$ here is slightly different from Theorem 5. In particular, $h : \alpha \mapsto (1 - \alpha)g(\alpha)$ is concave, but $h(1-) = \lim_{\alpha \to 1} \int_0^1 \text{VaR}_s(G) \, ds = 0$. Such $\rho$ may take an infinite value on $M_1$. As shown by Mao and Wang (2020, Theorem 3.1), any $\preceq_2$-consistent and translation invariant risk measure admits a representation as the infimum of benchmark-adjusted ES in this formulation.

The risk measures in (16) and (17) share visible similarity, and the main difference is that ES in (17) takes the place of VaR in (16). To foster a better understanding of these deep results, we roughly explain why these risk measures satisfy EMA for finite uncertainty sets (this relates to the “if” direction of both results), and more technical details are put in Appendix E.

(i) Let $\rho$ be the benchmark-loss VaR in (16). For any $F, G \in M_1$, suppose that $\rho(F) \leq 0$ and $\rho(G) \leq 0$. By definition of $\rho$, this means $F^{-1} \leq h$ and $G^{-1} \leq h$ and thus $\max\{F^{-1}, G^{-1}\} \leq h$, which in turn gives $\rho(\lor_1 \{F, G\}) \leq 0$. Further, using translation invariance gives EMA on $\{F, G\}$, which can be generalized to finite $\mathcal{F}$.

(ii) Let $\rho$ be the benchmark-adjusted ES in (17). For any $F, G \in M_1$, suppose that $\rho(F) \leq 0$ and $\rho(G) \leq 0$. Let $E_F(\alpha) = (1 - \alpha)\text{ES}_\alpha(F)$ and similarly for $E_G$. By definition of $\rho$, this
means $E_F \leq h$ and $E_G \leq h$ and thus $\max\{E_F, E_G\} \leq h$. Let $f^*$ be the concave envelope of a function $f$, that is, the infimum of concave functions dominating $f$. Since $h$ is concave, we have $(\max\{E_F, E_G\})^* \leq h^* = h$. Using $E_{\sqrt{2}\{F,G\}} = (\max\{E_F, E_G\})^*$ shown in Proposition EC.3, we get $\rho(\sqrt{2}\{F,G\}) \leq 0$. Further, using translation invariance gives EMA for finite $\mathcal{F}$. The above argument also hints at the reason why $h$ in Theorem 5 needs to be concave.

Different from Theorems 2, 3 and 4, lower semicontinuity is not assumed in Theorem 5. This point and some other technical points related to Theorem 5 are discussed in Appendix H.

**Example 3** (ES). For $\beta \in (0, 1)$, let $g(\alpha) = \infty \cdot 1_{\{\alpha > \beta\}}$, $\alpha \in [0, 1)$. The risk measure defined by (17) is $\rho = \text{ES}_\beta$. In Example 1, we have seen that ES$_\beta$ does not satisfy $\preceq_2$-EMA. Contrasting this observation with Theorem 5, we note that $h$ is not concave, and hence Theorem 5 does not apply.

**Example 4** (A new risk measure satisfying $\preceq_2$-EMA). Let $g(\alpha) = \alpha/(1 - \alpha)$, $\alpha \in [0, 1)$. The risk measure $\rho$ defined by (17) is real-valued on $\mathcal{M}_1$ and satisfies $\preceq_2$-EMA, which is given by

$$
\rho(F) = \sup_{\alpha \in [0, 1)} \left\{ \text{ES}_\alpha(F) - \frac{\alpha}{1 - \alpha} \right\}, \quad F \in \mathcal{M}_1.
$$

By Proposition 3.2 of Burzoni et al. (2022), the only positively homogeneous risk measures with the form (17) for some $g : [0, 1) \to [0, \infty]$ are ES$_\alpha$, $\alpha \in [0, 1)$. Since ES$_\alpha$ does not satisfy $\preceq_2$-EMA, there is no risk measure that satisfies translation invariance, $\preceq_2$-EMA, and positive homogeneity, contrasting the case of $\preceq_1$-EMA which leads to VaR in Corollary 1. We also note that there is no coherent risk measure satisfying $\preceq_1$-cEMA, $\preceq_1$-EMA or $\preceq_2$-EMA, whereas ES$_\alpha$ for $\alpha \in [0, 1)$ is the only class of coherent risk measures satisfying $\preceq_2$-cEMA.\footnote{For these statements on coherent risk measures, we do not assume the lower semicontinuity as in Theorems 2, 3 and 4. This is because any real-valued coherent risk measure is automatically $L^1$-continuous. Therefore, it suffices to verify that ES$_\alpha$ for $\alpha \in [0, 1)$ is the only class of risk measures satisfying all conditions in Theorem 3 with $L^1$-continuity in place of the lower semicontinuity, and no $L^1$-continuous coherent risk measures satisfy $\preceq_1$-cEMA via the same proof for Theorem 2. We omit the details here.}

**Remark 5.** EMA 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 $X \leq Y$ pointwisely. This leads to completely different interpretations and mathematics.

## 6 Comparing MA and WR in a few settings of uncertainty

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 WR and MA robust risk evaluation.
For results in this section and Section 7, 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

$$RVaR_{\alpha,\beta}(F) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} \text{VaR}_s(F) ds, \quad 0 \leq \alpha < \beta \leq 1.$$  

Special and limiting cases of RVaR$_{\alpha,\beta}$ include ES$_\alpha$ with $\beta = 1$ and VaR$_\beta$ with $\alpha \uparrow \beta$. If $\beta < 1$, then 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

$$PD_k(F) = \int_0^1 ks^{k-1} \text{VaR}_s(F) ds, \quad k \geq 1.$$  

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

$$\alpha \mathbb{E}[(X - t)_+] = (1 - \alpha) \mathbb{E}[(X - t)_-], \quad X \sim F \in M_1.$$  

The risk measure $\text{ex}_\alpha$ 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 $M_p$ be the set of distributions on $\mathbb{R}$ with finite $p$th moment and $F_0 \in M_p$ be a pre-specified distribution used as benchmark. For $p \geq 1$, the $\ell_p$-Wasserstein metric between $F$ and $F_0$ is defined as

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

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

$$\mathcal{F}_{p,\varepsilon}(F_0) = \{ F \in M_p : W_p(F, F_0) \leq \varepsilon \},$$  

which is a convex set. The parameter $\varepsilon$ represents the magnitude of uncertainty. Denote by

$$F_{p,\varepsilon|F_0}^1 = \bigvee_{1} \mathcal{F}_{p,\varepsilon}(F_0) \quad \text{and} \quad F_{p,\varepsilon|F_0}^2 = \bigvee_{2} \mathcal{F}_{p,\varepsilon}(F_0)$$
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 6.** Suppose that $\varepsilon > 0$, $p \geq 1$ and $F_0 \in \mathcal{M}_p$.

(a) The left quantile of $F^1_{p,\varepsilon|F_0}$ is given by the unique solution to the following equation

$$\left( \int_0^1 \left( (F^1_{p,\varepsilon|F_0})^{-1}(\alpha) - F^{-1}_0(s) \right)^p \frac{ds}{s} \right)^{\frac{1}{p}} = \varepsilon, \quad \alpha \in (0, 1).$$  \hspace{1cm} (20)

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

$$(F^2_{p,\varepsilon|F_0})^{-1}(\alpha) = F^{-1}_0(\alpha) + \left( 1 - \frac{1}{p} \right) (1 - \alpha)^{-\frac{1}{p}} \varepsilon, \quad \alpha \in (0, 1).$$ \hspace{1cm} (21)

Since the distribution functions $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$, as well as their quantile functions, are obtained explicitly in Theorem 6, the robust risk values $\rho^{\text{MA}}_{\preceq_1}(\mathcal{F}_{p,\varepsilon}(F_0))$ and $\rho^{\text{MA}}_{\preceq_2}(\mathcal{F}_{p,\varepsilon}(F_0))$ can be computed for any risk measure $\rho$ in a straightforward manner. On the other hand, $\rho^{\text{WR}}(\mathcal{F}_{p,\varepsilon}(F_0))$ is often difficult to compute if the risk measure is complicated, although there are some results in the literature considered the WR approach for special choices of risk measures; see Liu et al. (2022) for distortion risk measures, Blanchet et al. (2022) for mean-variance portfolio optimization, and Gao and Kleywegt (2016), Esfahani and Kuhn (2018) and Blanchet and Murthy (2019) for robust stochastic optimization.

As a feature of the robust model, both $F^1_{p,\varepsilon|F_0}$ and $F^2_{p,\varepsilon|F_0}$ are heavy-tailed distributions 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, $(F^1_{p,\varepsilon|F_0})^{-1} \geq (F^2_{p,\varepsilon|F_0})^{-1}$, and $(F^2_{p,\varepsilon|F_0})^{-1}$ is the sum of the quantile $F^{-1}_0$ and a Pareto quantile with tail index $p > 1$. Some other observations on the supremum distributions in Theorem 6 are made in Remark EC.8.

Noting that the Wasserstein uncertainty set $\mathcal{F}_{p,\varepsilon}(F_0)$ is convex, we have $\text{ES}^{\text{WR}}_{\alpha}(\mathcal{F}_{p,\varepsilon}(F_0)) = (\text{ES}^{\text{MA}}_{\alpha})_{\preceq_2}(\mathcal{F}_{p,\varepsilon}(F_0))$ by Theorem 1. A simulation result in case of $p = 2$, $\varepsilon = 0.1$ and a standard normal benchmark distribution is reported in Appendix F.2.

### 6.2 Multivariate Wasserstein uncertainty

For $p \geq 1$, let $\mathcal{M}_p(\mathbb{R}^d)$ be the set of all distributions on $\mathbb{R}^d$ with finite $p$th moment in each component. Let $\| \cdot \|$ be any given norm on $\mathbb{R}^d$ with its dual norm as $\| y \|_* = \sup_{\| x \| \leq 1} x^\top y$. For
example, $\| \cdot \|$ can be chosen as the $L^a$ norm for $a \geq 1$, and its dual norm is the $L^b$ norm such that $1/a + 1/b = 1$. The $\ell_p$-Wasserstein metric on $\mathbb{R}^d$ between $F,G \in \mathcal{M}_p(\mathbb{R}^d)$ is defined as

$$W^d_p(F,G) = \inf_{X \sim F, Y \sim G} \left( \mathbb{E}[\|X - Y\|^p] \right)^{1/p}.$$  

If $d = 1$, then $W^d_p$ is $W_p$ in (18) 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 (19),

$$\mathcal{F}^d_{p,\varepsilon}(F_0) = \left\{ F \in \mathcal{M}_p(\mathbb{R}^d) : W^d_p(F,F_0) \leq \varepsilon \right\}, \quad \varepsilon \geq 0.$$  

(22)

We focus on a portfolio selection problem, where the portfolio risk is $\rho(w^\top X)$ for some weight vector $w \in \mathbb{R}^d$ and risk vector $X$ with unknown distribution in the multi-dimensional Wasserstein ball $\mathcal{F}^d_{p,\varepsilon}(F_0)$. The univariate uncertainty set for the distribution of $w^\top X$ is denoted by

$$\mathcal{F}_{w,p,\varepsilon}(F_0) = \left\{ F^{w\top Z} : F \in \mathcal{F}^d_{p,\varepsilon}(F_0) \right\}, \quad F_0 \in \mathcal{M}_p(\mathbb{R}^d).$$  

(23)

We consider $p > 1$ since $\mathcal{F}^d_{1,\varepsilon}(F_0)$ is not bounded from above with respect to $\preceq_2$ as shown by Theorem 6. In the following theorem, we show that the problem of multivariate Wasserstein uncertainty can be conveniently converted to a univariate setting.

**Theorem 7.** For $\varepsilon \geq 0$ and $p > 1$, random vector $X$ with $F_X \in \mathcal{M}_p(\mathbb{R}^d)$ and $w \in \mathbb{R}^d$ with $w^\top w \neq 0$, we have $\mathcal{F}_{w,p,\varepsilon}(F_X) = \mathcal{F}_{p,\|w\|\varepsilon}(F^{w\top X})$. As a consequence, for any $\rho : \mathcal{M}_1 \to \mathbb{R}$ and $i \in \{1, 2\}$,

$$\rho^{WR}(\mathcal{F}_{w,p,\varepsilon}(F_X)) = \rho^{WR}(\mathcal{F}_{p,\|w\|\varepsilon}(F^{w\top X})) \quad \text{and} \quad \rho^{MA}_{\preceq i}(\mathcal{F}_{w,p,\varepsilon}(F_X)) = \rho^{MA}_{\preceq i}(\mathcal{F}_{p,\|w\|\varepsilon}(F^{w\top X})).$$

Intuitively, the dimension reduction result in Theorem 7 means that 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 risk value or the portfolio optimization explicitly by applying Theorem 6. In particular, the optimization problem under the $\preceq_2$-MA approach is much more tractable and can be computed efficiently, because Theorem 6 provides a direct and concise formula for the robust distribution with respect to $\preceq_2$. Although the WR approach admits the same dimension reduction technique, we note that it is not easy to compute for general risk measures even for the univariate Wasserstein uncertainty. The result in
Theorem 7 applies immediately to problems of the form $\rho(f(w^\top Z))$ for a real function $f$.

Below we present, in the special case of an elliptical benchmark distribution and a coherent distortion risk measure, the portfolio optimization problem via the WR and the $\preceq_2$-MA approaches can be written as a simple optimization problem. An elliptical distribution with characteristic generator $\psi$ is denoted by $E(\mu, \Sigma, \psi)$, which has normal and t-distributions as special cases; see McNeil et al. (2015, Chapter 6) for a precise definition. Let the benchmark distribution $F_0 = E(\mu, \Sigma, \psi)$ and denote by $F_\psi = E(0, 1, \psi)$. The risk measure $\rho$ is defined by $\rho_h(F) = \int_0^1 \text{VaR}_s(F) dh(s)$, where $h : [0, 1] \to [0, 1]$ is increasing and convex with $h(0) = 1 - h(1) = 0$. Noting that $\rho_h$ is translation invariant and positively homogeneous, the WR portfolio optimization problem is, by applying Proposition 4 of Liu et al. (2022) and Theorem 7,

$$\min_{w \in \mathcal{W}} : \rho_{WR, h}(F_0) = w^\top \mu + \rho_h(F_\psi) \sqrt{w^\top \Sigma w} + \xi(p, h)\epsilon \|w\|_s,$$

(24)

and the $\preceq_2$-MA portfolio optimization problem is, by applying Theorems 6 and 7,

$$\min_{w \in \mathcal{W}} : (\rho_{h})_{\preceq_2}^{MA}(F_0) = w^\top \mu + \rho_h(F_\psi) \sqrt{w^\top \Sigma w} + \xi(p, h)\epsilon \|w\|_s,$$

(25)

where $\mathcal{W}$ is the feasible set of $w$, and

$$\xi(p, h) = \left( \int_0^1 (h'_+(s))^{p/(p-1)} ds \right)^{(p-1)/p}, \quad \xi(p, h) = \frac{p-1}{p} \int_0^1 (1 - s)^{-1/p} dh(s).$$

In particular, (24) and (25) are second-order conic program when $\|\cdot\|$ is the $L^2$ norm which is denoted by $\|\cdot\|_2$; see e.g., Ben-Tal and Nemirovski (2001). Coherence of $\rho$ (convexity of $h$) is essential for the WR formula in (24) because general formulas are not available for non-convex distortions under Wasserstein uncertainty. In contrast, the MA formula (25) holds for any distortion risk measures (even if they may not be $\preceq_2$-consistent) which directly follows from Theorems 6 and 7. Numerical and empirical results on the above approaches for robust portfolio selection are presented in Section 7.2.

### 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

$$\mathcal{F}_{\mu, \sigma} := \{ F \in \mathcal{M}_2 : \mathbb{E}[F] = \mu \text{ and } \text{var}(F) = \sigma^2 \},$$

(26)
where $\mathbb{E}[F]$ and $\text{var}(F)$ represent the mean and the variance of $F$, respectively. The two equalities in (26) can be safely replaced by inequalities $\mathbb{E}[F] \leq \mu$ and $\text{var}(F) \leq \sigma^2$ 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., Ghaoui et al. (2003), Zhu and Fukushima (2009), Natarajan et al. (2010), Chen et al. (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 Ghaoui et al. (2003) and Corollary 1.1 of Jagannathan (1977) (see also Müller and Stoyan (2002, Theorem 1.10.7)) yield

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

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 $F_{\mu,\sigma}^1 = \bigvee_1 \mathcal{F}_{\mu,\sigma}$ and $F_{\mu,\sigma}^2 = \bigvee_2 \mathcal{F}_{\mu,\sigma}$ 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 $F_{\mu,\sigma}^1$ and $F_{\mu,\sigma}^2$.

**Proposition 3.** Suppose that $\mu \in \mathbb{R}$ and $\sigma > 0$. We have

$$
F_{\mu,\sigma}^1(x) = \frac{(x - \mu)^2}{\sigma^2 + (x - \mu)^2}, \quad x \geq \mu,
$$

and

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

We note that both $F_{\mu,\sigma}^1$ and $F_{\mu,\sigma}^2$ 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 $F_{\mu,\sigma}^1$ and $F_{\mu,\sigma}^2$ 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 $\mathcal{M}_1$.

By Proposition 3, 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 $F_{\mu,\sigma}^1$ or $F_{\mu,\sigma}^2$. To compute the WR robust risk value, for a coherent risk measure $\rho$, Li (2018) gives the explicit expression of $\rho^{\text{WR}}(\mathcal{F}_{\mu,\sigma})$ based on the Kusuoka representation. In addition, noting
Table 1: WR and MA under uncertainty induced by $\mathcal{F}_{0,1}$.

| $\rho$ | $\rho_{\text{WR}}$ | $\rho_{\text{MA}}^{\leq 1}$ | $\rho_{\text{MA}}^{\leq 2}$ |
|-------|------------------|----------------|----------------|
| ES$\alpha$ | $\frac{\alpha}{1-\alpha} \int_0^1 s^{\frac{1}{1-\alpha}} ds$ | $\frac{\alpha}{1-\alpha}$ | $\frac{\alpha}{1-\alpha}$ |
| RVaR$\alpha,\beta$ | $\frac{1}{\beta-\alpha} \int_0^\beta s^{\frac{1}{\beta-\alpha}} ds$ | - | - |
| VaR$\alpha$ | $\frac{1}{\beta-\alpha} \int_0^\beta s^{\frac{1}{\beta-\alpha}} ds$ | - | - |
| PD$\alpha$ | $\frac{k-1}{\sqrt{2k-1}} \frac{\sqrt{\pi} \Gamma(k+1/2)}{\Gamma(k)}$ | $\frac{\sqrt{\pi} \Gamma(k-1) \Gamma(k+1/2)}{2k-1}$ | $\frac{\sqrt{\Gamma(k)}}{\Gamma(k)}$ |
| ex$\alpha$ | $\frac{\alpha-1/2}{\sqrt{\alpha(1-\alpha)}}$ | $\text{ex}_{\alpha}(\mathcal{F}_{0,1})^1$ | $\frac{\alpha-1/2}{\sqrt{\alpha(1-\alpha)}}$ |

Note: $\Gamma$ is the gamma function; (RVaR$\alpha,\beta$)$_{\leq 2}^{\text{MA}}$ and (VaR$\alpha$)$_{\leq 2}^{\text{MA}}$ are not reported because RVaR$\alpha,\beta$ and VaR$\alpha$ are not $\preceq_2$-consistent; ex$_\alpha(\mathcal{F}_{0,1})$ can be numerically computed but it does not admit an explicit formula.

that $\mathcal{F}_{\mu,\sigma}$ is a convex set, if $\rho$ is an ES (Theorem 1) or benchmark-adjusted ES (Theorem 5), then $\rho_{\text{WR}}(\mathcal{F}_{\mu,\sigma}) = \rho_{\text{MA}}^{\leq 1}(\mathcal{F}_{\mu,\sigma}) = \rho(F_{1,\mu,\sigma}^-)$. If $\rho$ is a benchmark-loss VaR (Theorem 4), including the special case of VaR, then $\rho_{\text{WR}}(\mathcal{F}_{\mu,\sigma}) = \rho_{\text{MA}}^{\leq 1}(\mathcal{F}_{\mu,\sigma}) = \rho(F_{1,\mu,\sigma}^-)$. The explicit WR and MA robust risk values for ES$\alpha$, RVaR$\alpha,\beta$, the power-distorted risk measure and the expectile are given in Table 1,\textsuperscript{12} and a few figures on their numerical values are reported in Appendix F.2. Since those risk measure 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 $\rho(w^\top X)$ for some portfolio weight vector $w \in \mathbb{R}^d$ and risk vector $X$ with unknown distribution in the uncertainty set with given first two moments, which can be formulated as, for a feasible set $W$ of $w$,

$$\min_{w \in W} \rho_{\leq 1}^{\text{MA}}(\mathcal{F}_{w,\mu,\Sigma})$$, where $\mathcal{F}_{w,\mu,\Sigma} = \{F_{w\top X} : E[X] = \mu, \text{Cov}(X) = \Sigma\}$,

(29)

where $E[X]$ and Cov$(X)$ represents the mean vector and the covariance of $X$. Applying the general projection property in Popescu (2007) (see also Chen et al. (2011, Lemma 2.2)), the two sets $\mathcal{F}_{w,\mu,\Sigma}$ and $\mathcal{F}_{w\top \mu,\sqrt{w\top \Sigma w}}$ are identical. Hence, (29) is equivalent to

$$\min_{w \in W} \rho_{\leq 1}^{\text{MA}}(\mathcal{F}_{w\top \mu,\sqrt{w\top \Sigma w}})$$.

\textsuperscript{12}To obtain these formulas, we use the following results. Li et al. (2018) showed that RVaR$^{\text{WR}}_{\alpha,\beta}(\mathcal{F}_{\mu,\sigma}) = \text{ES}_{\alpha}^{\text{WR}}(\mathcal{F}_{\mu,\sigma})$ for all $\beta \in (0, 1)$. The value of PD$\alpha$ 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 3 and noting that all elements in $\mathcal{F}_{\mu,\sigma}$ have the same expectation, we obtain $\text{ex}_{\alpha}^{\text{WR}}(\mathcal{F}_{\mu,\sigma}) = (\text{ex}_{\alpha})_{\leq 2}^{\text{MA}}(\mathcal{F}_{\mu,\sigma})$. 

23
In case \( \preceq \) is \( \preceq_1 \) or \( \preceq_2 \), the above problem can be directly solved by applying Proposition 3. In particular, if \( \rho \) satisfies translation invariance and positive homogeneity, problem (30) leads to the following convenient formulation of second-order conic program, for \( i = 1, 2 \),

\[
\min_{w \in \mathcal{W}} \rho^M_{\preceq_i} \left( F_{w^\top \mu, \sqrt{w^\top \Sigma w}} \right) = \min_{w \in \mathcal{W}} \left\{ w^\top \mu + \sqrt{w^\top \Sigma w} \rho(F_{0,1}^i) \right\},
\]

where \( F_{0,1}^i \) is given by Proposition 3 in explicit form. Numerical and empirical results on the above MA approaches for robust portfolio selection is presented in Appendix G.

7 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 five stocks: Apple Inc. (AAPL), Amazon.com Inc. (AMZN), eBay Inc. (EBAY), Alphabet Inc. (GOOGL), Intel Corp. (INTC), where their historical loss data are collected from Yahoo! Finance. We use the period of January 1, 2019, to August 1, 2021, with a total of 649 observations of the daily losses of the five stocks.

We shall conduct two sets of numerical experiments. First, in Section 7.1, we present the robust distributions based on the MA approach when the uncertainty set consists of finite distributions 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 7.2, we consider the application of the MA approach with Wasserstein uncertainty as in Section 6.2, and data of all five stocks will be used.

7.1 Performance of MA with finite uncertainty set

We examine the MA approach for the uncertainty set consists of the distributions 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 distribution function is denoted by \( \hat{F} \). The uncertainty set \( \mathcal{F} \) consists of these four distributions, i.e., \( \mathcal{F} = \{ \hat{F}, F_n, F_t, F_{lg} \} \), and the supremum of \( \mathcal{F} \) with respect to \( \preceq_1 \) and \( \preceq_2 \) will be denoted by \( \bigvee_1 \mathcal{F} \) and \( \bigvee_2 \mathcal{F} \), respectively.

Figure 2 (top panels) shows the distribution functions and integrated survival functions defined by (1) of the elements in \( \mathcal{F} \). Noting that \( \bigvee_1 \mathcal{F}(x) = \inf \{ \hat{F}(x), F_n(x), F_t(x), F_{lg}(x) \} \) for \( x \in \mathbb{R} \), the supremum \( \bigvee_1 \mathcal{F} \) can be roughly divided into four parts. By Proposition 1, \( \bigvee_2 \mathcal{F} = F^* \in \mathcal{F} \) on \( (a, b) \) if \( F^* \) has 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: \( \bigvee_2 \mathcal{F} = F_n \) on \( (-\infty, 0.02) \);
Figure 2: Left: Distribution function; Middle: Integrated survival function; Right: Suprema of $F$.

\[V_2 F = \hat{F} \text{ on } [0.02, 0.0445); V_2 F = F_t \text{ on } [0.0445, \infty).\] The curves of $V_1 F$ and $V_2 F$ are given in Figure 2 (bottom panel) from which we can see that $V_2 F \preceq_1 V_1 F$. Moreover, $V_2 F$ has a jump at point 0.02 which can be explained by the difference between left and right derivatives of the integrated survival function of $V_2 F$ at point 0.02.

In the following, we compare the $\preceq_1$- and $\preceq_2$-MA robust risk values and the WR ones with the uncertainty set $F$. The risk measures are RVaR or ES. In the case of RVaR$\alpha,\beta$, we set $\alpha = 0.95$ and let $\beta$ range from 0.95 to 1. In the case of ES$\alpha$, $\alpha$ ranges from 0.9 to 1.

Figure 3 shows the value of RVaR$\alpha,\beta$ of the distributions in $F$, and RVaR$\alpha,\beta$ based on the MA and WR approaches, and Figure 4 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 3, 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 $F_n$ dominates other elements in $F$ on $[0.95, 0.9685]$ which implies that $(\text{RVaR}_{0.95,\beta})_{\preceq_1}^{\text{MA}}(F) = \text{RVaR}_{0.95,\beta}^{\text{WR}}(F) = \text{RVaR}_{0.95,\beta}(F_n)$ for $\beta \in [0.95, 0.9685]$. From Figure 4, we find that ES$\alpha(F_n)$ and ES$\alpha(F_{lg})$ are always smaller than ES$\alpha(\hat{F})$ and ES$\alpha(F_t)$ for $\alpha \in [0.9, 1]$. The reason is that financial market loss data are heavy-tailed empirically (see e.g., McNeil et al. (2015)), and ES with high level $\alpha$ focuses on the tail loss. In addition, the curve of ES$^{\text{MA}}$ always lies above the curve of ES$^{\text{WR}}$, which implies that the MA approach is more conservative.

### 7.2 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. A similar application with mean-variance uncertainty is presented in Appendix G.1. The MA approach will be contrasted to the WR approach and the standard sample average approximation (SAA) approach. We construct a portfolio from the five stocks
mentioned in the beginning of this section, whose daily losses are denoted by $X_1$ (AAPL), $X_2$ (AMZN), $X_3$ (EBAY), $X_4$ (GOOGL) and $X_5$ (INTC). The summary statistics of daily losses of the five stocks are given in Table 2. The wealth invested in the asset $X_i$ is denoted by $w_i$ for $i = 1, \ldots, 5$. Thus, the total loss from the investment of these five stocks is $w^\top X$, where $w = (w_1, \ldots, w_5)$ and $X = (X_1, \ldots, X_5)$. The feasible region of $w$ is the standard simplex $\Delta_5$.

We consider the setting in Section 6.2 where uncertainty is modeled by a multi-dimensional Wasserstein ball. A study with mean-variance uncertainty is reported in Appendix G.1 with similar findings. For the choice of the risk measure $\rho$, we will work with $\text{PD}_k$ defined in Section 6 to measure the portfolio risk. There are a few reasons for this choice. First, $\text{PD}_k$ is $\preceq_2$-consistent (which also implies $\preceq_1$-consistency). Second, the performance of WR approach and MA approach with $\preceq_2$ are the same in the portfolio optimization problem under the mean-variance or the Wasserstein uncertainty if the risk measure is selected as ES or expectile, so we move away from these two
classic choices. Third, the portfolio optimization problem of PD$_k$ leads to a convenient formulation of second-order conic program under the Wasserstein uncertainty as in Section 6.2.

As in many classic settings of portfolio selection, e.g., the classic framework of Markowitz (1952) where risk is measured by the variance, we assume that the investor has a target level of expected return rate and minimizes the risk. That is, with the constraint $E[w^\top X] \leq -r_0$ where $r_0$ is the expected return level, the investor minimizes $\rho(w^\top X)$. The expected return level $r_0$ takes values in $[0.0005, 0.0023]$ which allows for feasible efficient frontiers.

We set the parameter $p = 2$ and $\|\cdot\|_2$ in the Wasserstein uncertainty ball $\mathcal{F}_p,F_0$, 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 G.2. For the whole-period data, the fitted t-distribution has $\nu = 3.994$ degrees of freedom and its mean and correlation matrix are in Table 2. We apply the WR and the $\preceq_2$-MA approaches, and the corresponding portfolio optimization problems are converted to second-order conic programs which can be computed efficiently. By (24) and (25) in Section 6.2, the optimization problems via the WR and the MA approaches are, respectively,

$$\min_{w \in \Delta_5} : \rho^{WR} (\mathcal{F}_{w,2,\varepsilon}(F_0)) = w^\top \mu + PD_k(F_\nu)\sqrt{w^\top \Sigma w} + \zeta_k \varepsilon \sqrt{w^\top w} \quad \text{s.t.} \quad w^\top \mu \leq -r_0, \quad (31)$$

and

$$\min_{w \in \Delta_5} : \rho^{MA} (\mathcal{F}_{w,2,\varepsilon}(F_0)) = w^\top \mu + PD_k(F_\nu)\sqrt{w^\top \Sigma w} + \xi_k \varepsilon \sqrt{w^\top w} \quad \text{s.t.} \quad w^\top \mu \leq -r_0, \quad (32)$$
Figure 5: The optimized values of PD_k under Wasserstein uncertainty using the whole-period data. Left: r_0 = 0.0015, k = 10 and ε ∈ [0, 0.03]; Middle: r_0 = 0.0015, ε = 0.01 and k ∈ [1, 20]; Right: k = 10, ε = 0.01 and r_0 ∈ [0.0005, 0.0023].

where ζ_k = k/√(2k-1), ξ_k = (√πΓ(k+1))/(2Γ(k+1/2)), F_ν is the unit variance t-distribution with the tail parameter ν, and μ and Σ are the mean and the covariance of the fitted F_0. The SAA approach optimizes the portfolio according to the empirical distribution of the asset losses. Figure 5 presents the optimized robust risk values under Wasserstein uncertainty 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, 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 ε = 0.01 and let k and r vary. In practice, the parameter ε should not be too small; one may tune ε so that the empirical distribution 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 6 depicts the performance of the three approaches under the Wasserstein uncertainty over the remaining 300 trading days with r_0 = 0.0015 and ε = 0.01, and we set k = 2 (left) or k = 20 (right). In both cases, the MA and WR approaches, being robust methods, perform similarly (in Appendix G.1, we see that MA slightly outperforms WR using mean-variance uncertainty on the same data set). In the case k = 2, the SAA approach outperforms the other approaches in terms of
the realized wealth process. In the case $k = 20$, MA and WR both outperform the SAA approach after the first 150 trading days. Intuitively, this means that, during the period from Jan to Aug 2021, conservative investment strategies likely outperform non-conservative strategies. The similar performance of the MA and WR approaches is not a coincidence due to the similarity of problems (31) and (32) by noting that $\varepsilon$ is small.

Figure 6: Wealth evolution under Wasserstein uncertainty with $\varepsilon = 0.01$ and $r_0 = 0.0015$. Left: $k = 2$; Right: $k = 20$.

8 Concluding remarks and discussions

The MA approach for robust risk evaluation is proposed, along with a comprehensive study on its properties and implications. Below, we summarize some the advantages of the MA approach, which are illustrated and discussed through several technical results, in contrast to the WR approach.

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 value instead of the risk model (Section 3).

2. The MA approach does not depend on a specific risk measure but rather on a commonly used stochastic order. The robust model produced by the MA approach can be readily applied to different risk evaluation procedures and decision problems and can be used for calibration, analysis, and simulation, even without any risk measures (Section 3).
3. The MA robust risk value is easier to compute than the WR robust risk value in many practical situations, especially if $\mathcal{F}$ is non-convex (Section 4.1). In particular, in the case of ES, the MA approach provides great computational flexibility (Section 4.2). In some settings of uncertainty, the MA approach leads to explicit formulas for the robust model (Section 6), and it can easily handle Wasserstein uncertainty in portfolio selection (Section 6.2).

4. The MA approach is easier to optimize than the WR approach for a general risk measure as it does not require iterated computations of the risk measure for each distribution in the set of uncertainty. In particular, the case of optimizing ES is in general much more tractable via the MA approach than the WR approach (Section 4.3).

5. The MA approach gives rise to the useful property of EMA which characterizes VaR and ES (Section 5). 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. In particular, for convex sets of uncertainty (Theorems 1 and 3), ES is the unique class of coherent risk measures allowing for a conversion between the WR approach and the MA approach (with $\leq 2$) which is more tractable in optimization.

The MA approach requires a stochastic order to be specified. For an interpretation of prudent risk evaluation as in (5), the risk measure of interest should be consistent with this stochastic order. This does not seem to be a problem for practical applications, as commonly used (law-invariant) risk measures and decision criteria are at least monotone, and thus consistent with $\leq 1$. Moreover, all convex risk measures are consistent with respect to $\leq 2$. We recommend, in most applications, using $\leq 2$ in an MA approach as the default option, for its nice interpretation in decision theory (strong risk aversion) and tractable 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 distributions 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 distributions 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 distributions and they are naturally interpretable in this setting; third, many problems that are multivariate in nature often boil down to robust risk evaluation in one-dimension; see the setting in Section 4.3 and the problem of portfolio selection in Sections 6.2, 6.3 and 7.2. If desired by specific applications, the theory of the MA approach can be readily extended to multi-dimensional risk measures (see e.g., Embrechts and Puccetti (2006)) with the help from multivariate stochastic orders (see e.g., Shaked and Shanthikumar (2007)).

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 distribution as an element of equal importance ex ante, we can equip a prior probability measure $\mu$ 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

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

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

Bayraksan, G. and Love, D. K. (2015). Data-driven stochastic programming using phi-divergences. *The Operations Research Revolution* (pp. 1-19). INFORMS.

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.

Bignozzi, V., Burzoni, M. and Munari, C. (2020). Risk measures based on benchmark loss distributions. *Journal of Risk and Insurance, 87*(2), 437–475.

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

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

Cerreia-Vioglio, S., Hansen, L. P., Maccheroni, F. and Marinacci, M. (2021). Making decisions under model misspecification. *SSRN*: 3666424.

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.

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

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

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, forthcoming.*

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.

Filipović, D. and Svindland, G. (2012). The canonical model space for law-invariant convex risk measures is $L^1$. *Mathematical Finance, 22*(3), 585–589.

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

Gao, R. and Kleywegt, A. J. (2016). Distributionally robust stochastic optimization with Wasserstein distance. *arXiv preprint arXiv:1604.02199*.

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

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance, 14*(1),
Han, X., Wang, B., Wang, R. and Wu, Q. (2021). Risk concentration and the mean-Expected Shortfall criterion. *arXiv*: 2108.05066.

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.

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.

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.

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). Technical note: 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*, forthcoming.

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.

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

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
Millossovich, P., Tsanakas, A. and Wang, R. (2021). A theory of multivariate stress testing. SSRN: 3966204.
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.
Popescu, I. (2007). Robust mean-covariance solutions for stochastic optimization. Operations Research, 55(1), 98–112.
Quiggin, J. (1993). Generalized Expected Utility Theory: The Rank-dependent Model. Kluwer, the Netherlands.
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.
Rüschendorf, L. (2013). Mathematical Risk Analysis. Dependence, Risk Bounds, Optimal Allocations and Portfolios. Springer, Heidelberg.
Shaked, M. and Shanthikumar, J. G. (2007). Stochastic Orders. Springer Series in Statistics.
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.
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

A Setting and notation

We organize the appendices as follows. We first collect some extra notation and terminology in Appendix A. In Appendix B, we formally introduce the lattice theory and prove a generalized version of Proposition 1. The proofs of the main results, which are formulated on general domains, are presented in Appendices C (results in Sections 3 and 4), D (results in Section 5.1), E (results in Section 5.2) and F (results and omitted figures in Section 6). In Appendix G, we present numerical results for the MA approach in portfolio selection under mean-variance uncertainty and under Wasserstein uncertainty with a normal benchmark distribution, which complements the numerical studies in Section 7. Some technical remarks that were omitted from the main paper are collected in Appendix H.

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 $p$th moment, $p \in [0, \infty)$, and $L^\infty$ be the space of all bounded random variables. Accordingly, denote by $\mathcal{M}_p$, $p \in [0, \infty]$, the set of all distribution functions of random variables in $L^p$. On the set $\mathcal{M}_\infty$ of compactly supported distributions, we can define VaR$_0$, VaR$_1$ and ES$_1$ which are finite, by

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

and

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

Denote by $\mathcal{M}_{bb}$ the set of all distribution functions $F$ with a support bounded from below, i.e., $\text{VaR}_0(F) > -\infty$. In what follows, $\bigvee_1 \mathcal{F}$ and $\bigvee_2 \mathcal{F}$ represent the supremum of $\mathcal{F}$ in the ordered set $(\mathcal{M}_0, \preceq_1)$ and $(\mathcal{M}_1, \preceq_2)$, respectively. For two real objects (numbers or functions) $f$ and $g$, $f \vee g$ is their (point-wise) maximum, and $f \land g$ is their (point-wise) minimum.

A convex risk measure (Föllmer and Schied (2016)) is a risk measure which satisfies $\preceq_1$-consistency, translation invariance and convexity, i.e., relaxing positive homogeneity from a coherent risk measure. Recall that all risk measures in this paper are law-invariant.
B 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 \((M, \preceq)\) be an ordered set (i.e., \(\preceq\) is a partial order on \(M\)) and \(F \subseteq M\).

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

\[
U(F) = \{G \in M : F \preceq G, \forall F \in F\} \quad \text{and} \quad L(F) = \{G \in M : G \preceq F, \forall F \in F\}. \quad (EC.1)
\]

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

(iii) If for all \(F, G \in M\), \(\bigvee\{F, G\} \in M\) and \(\bigwedge\{F, G\} \in M\) for all \(F \subseteq M\) that is bounded from above and \(\bigwedge\{F, G\} \in M\) for all \(F \subseteq M\) that is bounded from below, then \((M, \preceq)\) is called a complete lattice.\(^{13}\)

Remark EC.1. In case \((M, \preceq)\) is a lattice which is not complete, \(\bigvee F\) may not exist even if \(F\) is bounded from above. In this case, the definition of the MA robust risk value needs to be modified. We can alternatively define \(\rho^{MA}_\preceq(F) = \inf_{F \in U(F)} \rho(F)\) where \(U(F) = \{G \in M : G \preceq F, \forall F \in F\}\), and this definition is equivalent to (4) if \((M, \preceq)\) is a complete lattice.

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) \(F(x) \geq G(x)\) for all \(x \in \mathbb{R}\); (iii) \(F^{-1}(\alpha) \leq G^{-1}(\alpha)\) for all \(\alpha \in (0, 1)\).

In case of \(\preceq_2\), the following statements are equivalent: (i) \(F \preceq_2 G\); (ii) \(\pi_F(x) \leq \pi_G(x)\) for all \(x \in \mathbb{R}\) where \(\pi_F\) is the integrated survival function defined by (1); (iii) \(E_F(\alpha) \leq E_G(\alpha)\) for all \(\alpha \in (0, 1)\) where \(E_F\) is the integrated quantile function defined by

\[
E_F(\alpha) = (1 - \alpha)ES_\alpha(F) = \int_0^1 F^{-1}(s)ds, \quad \alpha \in [0, 1]. \quad (EC.2)
\]

\(^{13}\) 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 \(M\) does not. Nevertheless, if we extend \(M\) to \(\overline{M} := M \cup \{F_{\text{min}}, F_{\text{max}}\}\) where \(F_{\text{min}} \preceq F\) and \(F \preceq F_{\text{max}}\) for all \(F \in M\), then our definition of complete lattice on the ordered set \((\overline{M}, \preceq)\) is equivalent to the one of Davey and Priestley (2002).
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 Rösl (2000). Here \(\mathcal{M}_+ = \{ F \in \mathcal{M}_0 : \int_0^\infty xdF(x) < \infty \}\). 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 ordered set \((\mathcal{M}_p, \preceq_1)\) is a complete lattice with the supremum \(\bigvee_1 \mathcal{F} = \inf_{F \in \mathcal{F}} F\) for \(\mathcal{F} \subseteq \mathcal{M}_p\) which is bounded from above. The left quantile function of \(\bigvee_1 \mathcal{F}\) is \(\sup_{F \in \mathcal{F}} F^{-1}\).

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

\[
\pi_{\bigvee_2 \mathcal{F}} = \sup_{F \in \mathcal{F}} \pi_F, \quad \bigvee_2 \mathcal{F} = 1 + \left(\sup_{F \in \mathcal{F}} \pi_F\right)'.
\]

(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 \(\bigvee_2 \{F, G\} = 1 + (\pi_F \lor \pi_G)'\) 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 \(F \leq H \leq G\), then \(H \in \mathcal{M}_p\). It suffices to verify that

1. \(0 \leq \lim_{x \to -\infty} H(x) \leq \lim_{x \to -\infty} G(x) = 0\) and \(1 \geq \lim_{x \to -\infty} H(x) \geq \lim_{x \to -\infty} F(x) = 1\), which imply \(H\) is a distribution, 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 \(\int_0^\infty x^p dH(x) \leq \int_0^\infty x^p dF(x) < \infty\) and \(\int_{-\infty}^0 (x)^p dH(x) \leq \int_{-\infty}^0 (x)^p dG(x) < \infty\). It follows that \(\int_{\mathbb{R}} |x|^p dH(x) < \infty\), that is, \(H \in \mathcal{M}_p\).

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 \(\mathcal{F} \subseteq \mathcal{M}_p\). If \(\mathcal{F}\) is bounded from above and define \(H = \inf_{F \in \mathcal{F}} F\) which is increasing and right-continuous, then there exists \(G \in \mathcal{M}_p\) such that \(F \succeq H \succeq G\) 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 \(G \succeq H \succeq F\) 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 Rösch (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, $\operatorname{sup}_{F \in \mathcal{F}} \pi_F(x) \leq \pi_G(x)$ for $x \in \mathbb{R}$. One can check that

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

2. $\lim_{x \to \infty} \pi_0(x) \leq \lim_{x \to \infty} \pi_G(x) = 0$ 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 $x + \pi_0(x) \geq x + \pi_F'(x)$ for all $X \in \mathbb{R}$. Noting that $\lim_{x \to -\infty} x + \pi_F'(x) = E[F^*] \in \mathbb{R}$, 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 $H = 1 + (\operatorname{sup}_{F \in \mathcal{F}} \pi_F)'_+$ is a distribution in $\mathcal{M}_1$. By definition of supremum, it is standard to check that $\bigvee_{\mathcal{F}} = H$.

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, $E_G(\alpha) \leq \inf_{F \in \mathcal{F}} E_F(\alpha)$ 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 distribution in $\mathcal{M}_1$, say $H$. By definition of infimum, it is standard to check $H = \bigwedge_{\mathcal{F}}$. It follows from the relation between a distribution 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 $\min\{F, G\} \leq F_1 \leq \max\{F, G\}$, and $F_2^{-1} = -(E_F \wedge E_G)'_-$ which implies $\min\{F^{-1}, G^{-1}\} \leq F_2^{-1} \leq \max\{F^{-1}, G^{-1}\}$, and hence, $\min\{F, G\} \leq F_2 \leq \max\{F, G\}$. 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 $x \leq −1$. We have $F \notin \mathcal{M}_p$ and for $y < −1$, let $F_y$ be a distribution 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 $F \notin M_p$, we have that $(M_p \preceq_2)$ is not a complete lattice.

\[ \text{C Proofs for results in Sections 3 and 4} \]

\textbf{Proof of Proposition 2.} For a fixed $x \in \mathbb{R}$, both $F \mapsto F(x)$ and $F \mapsto \pi_F(x)$ are linear in $F$. Hence, for $F \in \text{conv} \mathcal{F}$ with $F = \sum_{i=1}^{n} \lambda_i F_i$ where $(\lambda_1, \ldots, \lambda_n) \in \Delta_n$ and $F_i \in \mathcal{F}$ for $i = 1, \ldots, n$, there exist $G_1, G_2 \in \{F_1, \ldots, F_n\} \subseteq \mathcal{F}$ such that $G_1(x) \leq F(x)$ and $\pi_{G_2}(x) \geq \pi_F(x)$. The results follow immediately from Proposition 1.

\textbf{Proof of Theorem 1.} (a) Since $E[F] = \lim_{x \to -\infty} \{x + \pi_F(x)\}$ for each $F \in M_1$, we have

\[
E_{\preceq_2}^{MA}[\mathcal{F}] = E \left[ \bigvee_{2} \mathcal{F} \right] = \lim_{x \to -\infty} \left\{ x + \sup_{F \in \mathcal{F}} \pi_F(x) \right\}
\]

\[
= \lim_{x \to -\infty} \sup_{F \in \mathcal{F}} \left\{ E[F] + \int_{\mathbb{R}} (x-y)_+ dF(y) \right\}
\]

\[
\leq \sup_{F \in \mathcal{F}} E[F] + \lim_{x \to -\infty} \sup_{F \in \mathcal{F}} \int_{\mathbb{R}} (x-y)_+ dF(y)
\]

\[
= \sup_{F \in \mathcal{F}} E[F] = E^{WR}[\mathcal{F}],
\]

The converse direction $E_{\preceq_2}^{MA}[\mathcal{F}] \geq E^{WR}[\mathcal{F}]$ is trivial. Hence, we complete the proof of (a).

(b) Suppose that $\mathcal{F} \subseteq M_1$ is a convex set which is bounded from above with respect to $\preceq_2$. Denote by $\Pi_G = \sup_{F \in \mathcal{G}} \pi_F$ for any set $\mathcal{G} \subseteq M_1$. If $\mathcal{G}$ is a convex polytope, then by Theorem 1 of Zhu and Fukushima (2009), we have

\[
E_{\alpha}^{WR}(\mathcal{G}) = (E_{\alpha})_{\preceq_2}^{MA}(\mathcal{G}). \tag{EC.3}
\]

Let $c = (E_{\alpha})_{\preceq_2}^{MA}(\mathcal{F})$. Using (8), we get

\[
x + \frac{1}{1-\alpha} \Pi_F(x) \geq c \text{ for all } x \in \mathbb{R}. \tag{EC.4}
\]

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

\[
x + \frac{1}{1-\alpha} \pi_G(x) \geq c \text{ for all } x < x_0. \tag{EC.5}
\]
Fix $\varepsilon > 0$. Let $G \subseteq F$ be a convex polytope such that

$$\Pi_G(x) \geq \Pi_F(x) - \varepsilon \text{ for all } x \in [x_0, c].$$ (EC.6)

Such a set $G$ exists since $\Pi_F$ is a decreasing convex function, which can be uniformly approximated by a discrete grid on the compact set $[x_0, c]$. Let $G_0 = \text{conv}(G \cup \{G\}) \subseteq F$, which is again a convex polytope. Using (EC.3), (EC.4), (EC.5) and (EC.6), we obtain

$$\text{ES}_WR^\alpha(G_0) = (\text{ES}_\alpha)_{\leq 2}^\text{MA}(G_0)$$

$$= \min \left\{ x + \frac{\Pi_{G_0}(x)}{1 - \alpha} \right\}$$

$$= \min \left\{ \inf_{x < x_0} \left\{ x + \frac{\Pi_{G_0}(x)}{1 - \alpha} \right\}, \min_{x \in [x_0, c]} \left\{ x + \frac{\Pi_{G_0}(x)}{1 - \alpha} \right\}, \inf_{x > c} \left\{ x + \frac{\Pi_{G_0}(x)}{1 - \alpha} \right\} \right\}$$

$$\geq \min \left\{ \inf_{x < x_0} \left\{ x + \frac{\pi_{G}(x)}{1 - \alpha} \right\}, \min_{x \in [x_0, c]} \left\{ x + \frac{\Pi_{G}(x)}{1 - \alpha} \right\}, c \right\}$$

$$\geq \min \left\{ \min_{x \in [x_0, c]} \left\{ x + \frac{\Pi_{F}(x)}{1 - \alpha} \right\}, c \right\} - \frac{\varepsilon}{1 - \alpha} = c - \frac{\varepsilon}{1 - \alpha}.$$

Note that $\text{ES}_WR^\alpha(F) \geq \text{ES}_WR^\alpha(G_0) \geq c - \varepsilon/(1 - \alpha)$ because $G_0 \subseteq F$. Since $\varepsilon$ is arbitrary, we get $\text{ES}_WR^\alpha(F) \geq c = (\text{ES}_\alpha)_{\leq 2}^\text{MA}(F)$. Together with $\text{ES}_WR^\alpha(F) \leq (\text{ES}_\alpha)_{\leq 2}^\text{MA}(F)$, we obtain the desired equality $\text{ES}_WR^\alpha(F) = (\text{ES}_\alpha)_{\leq 2}^\text{MA}(F)$.

(c) It follows directly from Proposition 1.

D Proofs and generalizations for results in Section 5.1

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

$$\rho \left( \bigvee_i \{F_1, \ldots, F_n\} \right) = \sup \left\{ \rho \left( \sum_{i=1}^n \lambda_i F_i \right) : (\lambda_1, \ldots, \lambda_n) \in \Delta_n \right\}$$

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

D.1 A generalization of Theorem 2 and related results

The following theorem is a generalized version of Theorem 2 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 $\rho = \text{VaR}_\alpha$ for some $\alpha \in (0, 1)$.

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

**Lemma EC.1.** 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 $F \in \mathcal{M}_{bb} \cap \mathcal{M}_p$.

**Proof.** For $F \in \mathcal{M}_{bb} \cap \mathcal{M}_p$, there exists a sequence $\{F_n\}_{n \in \mathbb{N}} \subseteq \mathcal{M}_\infty$ such that $F_n \preceq_1 F_{n+1}$ for all $n \geq 1$ and $F_n \overset{d}{\to} F$. We have

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

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

Denote by $\ell$ the Lebesgue measure on $[0, 1]$ and $\mathcal{M}_{1,f}(\ell)$ the space of all finitely additive probability measures on $([0, 1], \mathcal{B}([0, 1]))$ that are absolutely continuous with respect to $\ell$. By Theorem 4.5 of Jia et al. (2020), for any $\preceq_1$-consistent, translation invariant and positively homogeneous risk measure $\rho : \mathcal{M}_\infty \to \mathbb{R}$, there exists a family $\{\mathcal{M}_\xi : \xi \in \Xi\}$ of nonempty, weak$^*$-compact and convex subsets of $\mathcal{M}_{1,f}(\ell)$ such that

$$\rho(F) = \min_{\xi \in \Xi} \max_{\mu \in \mathcal{M}_\xi} \int_0^1 \text{VaR}_s(F) \mu(ds), \quad F \in \mathcal{M}_\infty. \quad \text{(EC.7)}$$

Applying this representation, we can establish the following lemma.

**Lemma EC.2.** 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 $\bar{\rho}$ the constraint of $\rho$ on $\mathcal{M}_\infty$, i.e., $\bar{\rho}(F) = \rho(F)$ for all $F \in \mathcal{M}_\infty$. Then the following three statements are equivalent.

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

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

(c) $\bar{\rho}$ admits a representation (EC.7) which satisfies $\max_{\xi \in \Xi} \min_{\mu \in \mathcal{M}_\xi} \mu([0, s]) = 1_{\{s \geq \alpha\}}$ for some $\alpha \in (0, 1)$. 41
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 \( \{ M_\xi : \xi \in \Xi \} \) consisting of nonempty, weak*-compact and convex subsets of \( M_1, f(\ell) \) such that \( \tilde{\rho} \) admits a representation (EC.7). Denote by

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

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 \( g_\Xi \). Note that for any \( \beta \in [0, 1) \),

\[
\rho(\beta \delta_0 + (1 - \beta) \delta_1) = \min_{\xi \in \Xi} \max_{\mu \in M_\xi} \int_0^1 \VaR_s(\beta \delta_0 + (1 - \beta) \delta_1) \mu(ds) \\
= \min_{\xi \in \Xi} \max_{\mu \in M_\xi} \int_{[\beta, 1]} 1(\mu(ds) \\
= \min_{\xi \in \Xi} \max_{\mu \in M_\xi} \mu((\beta, 1]) = \min_{\xi \in \Xi} \max_{\mu \in M_\xi} (1 - g_\mu(\beta)) = 1 - g_\Xi(\beta).
\]

(EC.8) 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 \( \beta_n \delta_0 + (1 - \beta_n) \delta_1 \leq_1 \beta \delta_0 + (1 - \beta) \delta_1 \), \( n \in \mathbb{N} \), and \( \beta_n \delta_0 + (1 - \beta_n) \delta_1 \xrightarrow{d} \beta \delta_0 + (1 - \beta) \delta_1 \) as \( n \to \infty \). Therefore, by (EC.8), we have

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

where the inequality comes from \( \beta_n \delta_0 + (1 - \beta_n) \delta_1 \leq_1 \beta \delta_0 + (1 - \beta) \delta_1 \). By lower semicontinuity of \( \rho \), we have

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

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

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 \( g_\Xi \), we obtain that for all \( \xi \in \Xi \), there exists \( \mu \in M_\xi \) such that \( g_\mu(1-) \leq g_\Xi(1-) = 1 - \delta \). Take \( G \in M_p \setminus M_\infty \), \( \rho \in [0, \infty) \) with support on \( \mathbb{R}_+ \). For any \( M \in \mathbb{R}_+ \), there exists \( \beta \in [0, 1) \) such that \( F := \beta \delta_0 + (1 - \beta) \delta_1 \in M_\infty \) and \( F \leq_1 G \). Therefore,

\[
\rho(G) \geq \rho(F) = \tilde{\rho}(F) = \min_{\xi \in \Xi} \max_{\mu \in M_\xi} \int_0^1 \VaR_s(F) \mu(ds) = M(1 - g_\Xi(\beta)) \geq \delta M.
\]

42
Since $M$ is arbitrary, this yields a contradiction to that $\rho : M_p \to \mathbb{R}$, $p \in [0, \infty)$.

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

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

(EC.9)

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 $F = \beta_0 + (1 - \beta)\delta_1$ and $G = \delta_0$ with $\beta \in (0, 1)$ and $a > 0$. We can calculate the MA robust risk value

$$\tilde{\rho} \left( \bigvee_{0}^{1} \{F, G\} \right) = \tilde{\rho}(\beta_0 + (1 - \beta)\delta_1) = 1 - g_\Xi(\beta),$$

and the WR robust risk value

$$\sup_{\lambda \in [0, 1]} \tilde{\rho}(\lambda F + (1 - \lambda)G) = \sup_{\lambda \in [0, 1]} \tilde{\rho}(\lambda\beta_0 + (1 - \lambda)\delta_0 + \lambda(1 - \beta)\delta_1)$$

$$= \sup_{\lambda \in [0, 1]} \min_{\xi \in \Xi} \max_{\mu \in M_\xi} \{-ag_\mu(\lambda\beta) + 1 - g_\mu(1 - \lambda)\}$$

$$\leq 1 - \inf_{\lambda \in [0, 1]} \max_{\xi \in \Xi} \{a \min_{\mu \in M_\xi} g_\mu(\lambda\beta) + \min_{\mu \in M_\xi} g_\mu(1 - \lambda)\}$$

$$= 1 - \inf_{\lambda \in [0, 1]} \max_{\xi \in \Xi} \{a g_\xi(\lambda\beta) + g_\xi(1 - \lambda)\}.$$ 

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

$$g_\Xi(\beta) \geq \inf_{\lambda \in [0, 1]} \max_{\xi \in \Xi} \{a g_\xi(\lambda\beta) + g_\xi(1 - \lambda)\}, \quad \beta \in [0, 1], \quad a > 0.$$ (EC.10)

Fix $\beta > 0$. By (EC.10) and the definition of infimum, for each $a > 0$, there exist $\varepsilon_a$ and $\lambda_a \in [0, 1]$ such that $\lim_{a \to \infty} \varepsilon_a = 0$ and

$$g_\Xi(\beta) \geq \max_{\xi \in \Xi} \{a g_\xi(\lambda_a\beta) + g_\xi(1 - \lambda_a)\} - \varepsilon_a.$$ (EC.11)

It follows that $g_\Xi(\beta) \geq \max_{\xi \in \Xi} a g_\xi(\lambda_a\beta) - \varepsilon_a = a g_\Xi(\lambda_a\beta) - \varepsilon_a$ 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.11) implies $g_\Xi(\beta) \geq \max_{\xi \in \Xi} g_\xi(1 - \lambda_a) - \varepsilon_a = g_\Xi(1 - \lambda_a) - \varepsilon_a$. Letting $a \to \infty$, we have $g_\Xi(\beta) \geq g_\Xi(1-) = 1$ 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 $\alpha$ in (EC.9), we have $g_\Xi(\beta) = 0$ for all $\beta \in [0,\alpha)$. Fix $\beta > \alpha$, there exist $\varepsilon_\alpha$ and $\lambda_\alpha \in [0,1]$ such that $\lim_{\varepsilon_\alpha \to \infty} \varepsilon_\alpha = 0$ and (EC.11) holds. Similarly, we have $g_\Xi(\lambda_\alpha \beta) \to 0$ as $\alpha \to \infty$ which implies $\limsup_{\varepsilon_\alpha \to \infty} \varepsilon_\alpha \leq \alpha/\beta$. It then follows that $g_\Xi(\beta) \geq g_\Xi(1 - \lambda_\alpha(1 - \beta)) - \varepsilon_\alpha$. Letting $\alpha \to \infty$, we have

$$g_\Xi(\beta) \geq \limsup_{\alpha \to \infty} g_\Xi(1 - \lambda_\alpha(1 - \beta)) \geq g_\Xi \left( \frac{1 - \alpha(1 - \beta)}{\beta} \right) - \varepsilon_\alpha.$$

Since $\beta < 1 - \alpha(1 - \beta)/\beta$ for $\beta > \alpha$, and the sequence $\{\beta_n\}_{n \in \mathbb{N}}$, where $\beta_0 = \beta$ and $\beta_{n+1} = 1 - \alpha(1 - \beta)/\beta_n$ for $n \geq 0$, converges to 1, and $g_\Xi$ 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 $g_\Xi$, we have $g_\Xi(\beta) = \mathds{1}_{\{\beta \geq \alpha\}}$, which completes the proof of (b)$\Rightarrow$(c).

(c)$\Rightarrow$(a): By (EC.8), under the condition of (c), for $F = \beta \delta_0 + (1 - \beta)\delta_1$, we have $\rho(F) = 1 - g_\Xi(\beta) = \text{VaR}_\alpha(F)$. By positive homogeneity and translation invariance of $\rho$, this implies that for any $F = \beta \delta_0 + (1 - \beta)\delta_1$, $\rho(F) = \text{VaR}_\alpha(F)$. For $F \in \mathcal{M}_\infty$, define $G = \beta \delta_{\text{VaR}_\alpha(F)} + (1 - \beta)\delta_{\text{VaR}_\beta(F)}$ for $\beta < \alpha$. One can check $G \preceq_1 F$ and thus, $\rho(F) \geq \rho(G) = \text{VaR}_\alpha(G) = \text{VaR}_\beta(F)$ for $\beta < \alpha$. By left-continuity of VaR, we have $\rho(F) \geq \lim_{\beta \uparrow \alpha} \text{VaR}_\beta(F) = \text{VaR}_\alpha(F)$. On the other hand, define $H = \alpha \delta_{\text{VaR}_\alpha(F)} + (1 - \alpha)\delta_{\text{VaR}_1(F)}$. One can check $F \preceq_1 H$ and thus, $\rho(F) \leq \rho(H) = \text{VaR}_\alpha(H) = \text{VaR}_\alpha(F)$. Therefore, we have $\rho(F) = \text{VaR}_\alpha(F)$. 

\[\quad\] 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.2, we have $\rho = \text{VaR}_\alpha$ for some $\alpha \in (0,1)$ on $\mathcal{M}_\infty$. It remains to show that $\rho = \text{VaR}_\alpha$ on $\mathcal{M}_p$, $p \in [0,\infty)$. By Lemma EC.1, we have $\rho = \text{VaR}_\alpha$ on $\mathcal{M}_{bb} \cap \mathcal{M}_p$. Next, we will prove $\rho = \text{VaR}_\alpha$ on $\mathcal{M}_p \setminus \mathcal{M}_{bb}$. Take $F \in \mathcal{M}_p \setminus \mathcal{M}_{bb}$ and $\zeta < \text{VaR}_\alpha(F)$. We have

$$\rho \left( \bigvee_1 \{F, \delta_\zeta\} \right) = \text{VaR}_\alpha \left( \bigvee_1 \{F, \delta_\zeta\} \right) = \text{VaR}_\alpha(F) \vee \zeta = \text{VaR}_\alpha(F),$$

where the first equality is due to $\bigvee_1 \{F, \delta_\zeta\} \in \mathcal{M}_{bb} \cap \mathcal{M}_p$ 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.12) implies that

$$\rho(F) \leq \text{VaR}_\alpha(F) \text{ for all } F \in \mathcal{M}_p \setminus \mathcal{M}_{bb}.$$
By $\preceq_1$-cEMA of $\rho$, and together with (EC.12), we have

$$\sup_{0 \leq \lambda \leq 1} \rho(\lambda F + (1 - \lambda)\delta_\zeta) = \text{VaR}_\alpha(F) \text{ for } \zeta < \text{VaR}_\alpha(F). \quad (\text{EC.14})$$

We assert that if $F$ is continuous at $\text{VaR}_\alpha(F)$, then (EC.14) implies

$$\limsup_{\lambda \to 1} \rho(\lambda F + (1 - \lambda)\delta_\zeta) = \text{VaR}_\alpha(F), \quad \zeta < \text{VaR}_\alpha(F). \quad (\text{EC.15})$$

To see this, fix $\varepsilon > 0$ and $\zeta < \text{VaR}_\alpha(F)$, and let $\lambda \in [0, 1 - \varepsilon)$. Since $F$ is continuous at $\text{VaR}_\alpha(F)$, we have

$$\zeta' := \text{VaR}_{(\alpha - \varepsilon)/(1 - \varepsilon)}(F) < \text{VaR}_\alpha(F).$$

Hence, the distribution $\lambda F + (1 - \lambda)\delta_\zeta$ has probability at least $\alpha - \varepsilon + \varepsilon \geq \alpha$ for the interval $(-\infty, \zeta' \lor \zeta]$. Therefore, we have $\text{VaR}_\alpha(\lambda F + (1 - \lambda)\delta_\zeta) \leq \zeta' \lor \zeta < \text{VaR}_\alpha(F)$. Hence,

$$\sup_{0 \leq \lambda \leq 1 - \varepsilon} \rho(\lambda F + (1 - \lambda)\delta_\zeta) \leq \zeta' \lor \zeta < \text{VaR}_\alpha(F).$$

Therefore, the supremum in (EC.14) is not attained on $[0, 1 - \varepsilon)$ for any $\varepsilon > 0$, and we have

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

By lower semicontinuity of $\rho$, we have

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

Combining above two equations, the assertion (EC.15) is verified. In the following, we will show that $\rho(G) = \text{VaR}_\alpha(G)$ for $G \in \mathcal{M}_p \setminus \mathcal{Mb}$ in two cases by applying (EC.13) and (EC.15).

**Case 1:** $G$ is continuous at $\text{VaR}_\alpha(G)$. By (EC.13), $\rho(G) \leq \text{VaR}_\alpha(G)$, and we suppose by contradiction that $\rho(G) < \text{VaR}_\alpha(G)$. Since $G$ is continuous at $\text{VaR}_\alpha(G)$, there exist $x^* \in (\rho(G), \text{VaR}_\alpha(G))$ such that $G$ is continuous at $x^*$, and $x^*$ is an element of the support of $G$, which implies $\text{VaR}_{G(x^*)}(G) = x^*$. Noting that $x^* < \text{VaR}_\alpha(G)$, we have $\lambda^* := G(x^*)/\alpha \in (0, 1)$. Define a distribution function as

$$H(x) = \begin{cases} \frac{1}{\lambda^*} G(x) \land 1, & x < \text{VaR}_\alpha(G) \\ 1, & x \geq \text{VaR}_\alpha(G). \end{cases}$$
One can check that $H$ is continuous at $x^*$ and $\text{VaR}_\alpha(H) = x^*$. By (EC.15), we obtain for $\zeta < x^*$,

$$\limsup_{\lambda \to 1} \rho(\lambda H + (1 - \lambda)\delta_\zeta) = \text{VaR}_\alpha(H) = x^*. \quad (\text{EC.16})$$

For $\zeta < x^*$ and $\lambda \in [\lambda^*, 1]$, we have $\lambda H + (1 - \lambda)\delta_\zeta \succeq G$ pointwisely, which implies $\lambda H + (1 - \lambda)\delta_\zeta \preceq G$. It follows from $\leq_1$-consistency that $\rho(G) \succeq \rho(\lambda H + (1 - \lambda)\delta_\zeta)$ for all $\zeta < x^*$ and $\lambda \in [\lambda^*, 1]$. Hence, by (EC.16), we obtain $\rho(G) \geq x^*$, and this yields a contradiction.

**Case 2:** $G$ has a jump at $\text{VaR}_\alpha(G)$. In this case, we can construct a sequence $\{G_n\}_{n \in \mathbb{N}}$ such that $G_n \preceq F$ for all $n \in \mathbb{N}$, $G_n$ is continuous at point $\text{VaR}_\alpha(G_n)$ and $\text{VaR}_\alpha(G_n) \to \text{VaR}_\alpha(G)$. By Case 1 and $\preceq_1$-consistency of $\rho$, we have $\rho(G) \geq \rho(G_n) \to \text{VaR}_\alpha(G)$. Note that the converse direction holds by (EC.13). Hence, we obtain $\rho(G) = \text{VaR}_\alpha(G)$ for all $F \in \mathcal{M}_p \setminus \mathcal{M}_{bb}$ such that $G$ has a jump at point $\zeta$.

In summary, we complete the proof of this theorem.

**D.2 A generalization of Theorem 3 and related results**

The following theorem is a generalized result of Theorem 3 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 $\rho = \text{ES}_\alpha$ 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

$$\rho(F) = \min_{\xi \in \Xi} \max_{h \in \mathcal{H}_\xi} \int_0^1 \text{VaR}_\alpha(F)dh(s), \ F \in \mathcal{M}_\infty. \quad (\text{EC.17})$$

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 $\rho(F) = \int_0^1 \text{VaR}_\alpha(F)dh(s)$. Therefore, for any $\beta \in [0, 1]$, by $\rho(\beta\delta_0 + (1 - \beta)\delta_1) = \min_{\xi \in \Xi} \max_{h \in \mathcal{H}_\xi} \int_0^1 1dh(s) = 1 - \max_{\xi \in \Xi} \min_{h \in \mathcal{H}_\xi} h(\beta)$, we can define

$$h_\xi(\beta) = \min_{h \in \mathcal{H}_\xi} h(\beta) \quad \text{and} \quad h(\beta) = \max_{\xi \in \Xi} h_\xi(\beta), \ \beta \in [0, 1]. \quad (\text{EC.18})$$
Applying this representation, we can establish the following lemma.

**Lemma EC.3.** Let \( \rho : M_p \to \mathbb{R}, p \in [1, \infty) \), be a mapping satisfying \( \preceq_2 \)-consistency, translation invariance and positive homogeneity. Denote by \( \bar{\rho} \) the constraint of \( \rho \) on \( M_\infty \). Then the following three statements are equivalent.

(a) \( \bar{\rho} = ES_\alpha \) for some \( \alpha \in [0,1) \) on \( M_\infty \).

(b) \( \bar{\rho} \) satisfies \( \preceq_2 \)-cEMA.

(c) \( \bar{\rho} \) admits a representation (EC.17) which satisfies \( h_\Xi(\beta) = (\beta - \alpha)_+/(1 - \alpha) \) for some \( \alpha \in [0,1) \) where \( h_\Xi \) is defined by (EC.18).

**Proof.** The implication (a) \( \Rightarrow \) (b) follows immediately from Theorem 1 (b).

\( (b) \Rightarrow (c) \): By the discussion before this lemma, there exists a family \( \{ H_\xi : \xi \in \Xi \} \) of nonempty, compact and convex subsets of \( H \) such that \( \bar{\rho} \) admits a representation (EC.17). Define \( h_\xi \) and \( h_\Xi \) by (EC.18). One can check that both \( h_\xi \) and \( h_\Xi \) 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 a similar arguments in (b) \( \Rightarrow \) (c) of Lemma EC.2.

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

\[
\bar{\rho}\left(\bigvee_2 \{F,G\}\right) = \bar{\rho}\left(\beta \delta_{-a} + (1 - \beta) \delta_1\right) = 1 - \frac{h_\Xi(\beta)}{\beta},
\]

and the WR robust risk value

\[
\sup_{\lambda \in [0,1]} \bar{\rho}(\lambda F + (1 - \lambda)G) = \bar{\rho}(\lambda \beta \delta_{-a} + (1 - \lambda) \delta_0 + \lambda(1 - \beta) \delta_1)
\]

\[
= \sup_{\lambda \in [0,1]} \min_{\xi \in \Xi} \max_{h \in H_\xi} \{a h(\lambda \beta) + 1 - \lambda(1 - \beta)\}
\]

\[
= 1 - \inf_{\lambda \in [0,1]} \max_{\xi \in \Xi} \min_{h \in H_\xi} \{a h(\lambda \beta) + h(1 - \lambda(1 - \beta))\}
\]

\[
\leq 1 - \inf_{\lambda \in [0,1]} \max_{\xi \in \Xi} \min_{h \in H_\xi} \{a h(\lambda \beta) + h(1 - \lambda(1 - \beta))\}
\]

\[
= 1 - \inf_{\lambda \in [0,1]} \max_{\xi \in \Xi} \{a h_\xi(\lambda \beta) + h_\xi(1 - \lambda(1 - \beta))\}.
\]

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

\[
\frac{h_\Xi(\beta)}{\beta} \geq \inf_{\lambda \in [0,1]} \max_{\xi \in \Xi} \{a h_\xi(\lambda \beta) + h_\xi(1 - \lambda(1 - \beta))\}, \quad \beta \in [0,1], \quad a > \frac{1 - \beta}{\beta}. \tag{EC.19}
\]
Define

$$
\alpha = \inf \{ \beta : h_\varepsilon(\beta) > 0 \} \in [0, 1), \tag{EC.20}
$$

where the fact $\alpha < 1$ comes from $h_\varepsilon(1^-) = 1$. Fix $\beta > \alpha$. By (EC.19) and the definition of infimum, for each $a > (1 - \beta)/\beta$, there exist $\lambda_a \in [0, 1]$ and $\varepsilon_a$ such that $\lim_{a \to \infty} \varepsilon_a = 0$ and

$$
\frac{h_\varepsilon(\beta)}{\beta} \geq \max_{\xi \in \Xi} \{ ah\xi(\lambda_a \beta) + h\xi(1 - \lambda_a(1 - \beta)) \} - \varepsilon_a. \tag{EC.21}
$$

It follows that $h_\varepsilon(\beta)/\beta \geq \max_{\xi \in \Xi} a h\xi(\lambda_a \beta) - \varepsilon_a = a h\xi(\lambda_a \beta) - \varepsilon_a$. Letting $a \to \infty$, we obtain $\lim_{a \to \infty} h_\varepsilon(\lambda_a \beta) = 0$. By definition of $\alpha$, $\limsup_{a \to \infty} \lambda_a \leq \alpha/\beta$. Again, by (EC.21), we have

$$
\frac{h_\varepsilon(\beta)}{\beta} \geq \max_{\xi \in \Xi} \{ h\xi(1 - \lambda_a(1 - \beta)) \} - \varepsilon_a = h\xi(1 - \lambda_a(1 - \beta)) - \varepsilon_a \tag{EC.22}
$$

By monotonicity of $h_\varepsilon$ and $\limsup_{a \to \infty} \lambda_a \leq \alpha/\beta$, we get $\limsup_{a \to \infty} h_\varepsilon(1 - \lambda_a(1 - \beta)) - \varepsilon_a \geq h\xi((1 - \alpha(1 - \beta)/\beta) - \varepsilon_a$. This combined with (EC.22) implies $h_\varepsilon(\beta)/\beta \geq h\xi((1 - \alpha(1 - \beta)/\beta) - \varepsilon_a$. Denote by $h_\varepsilon(x) = \lim_{y \uparrow x} h_\varepsilon(y)$. We have $h_\varepsilon(\alpha) = 0$ and

$$
\frac{h_\varepsilon(\beta) - h_\varepsilon(\alpha)}{\beta - \alpha} \geq \frac{h_\varepsilon(\alpha + (1 - \alpha/\beta)) - h_\varepsilon(\alpha)}{1 - \alpha/\beta}, \quad \beta > \alpha. \tag{EC.23}
$$

Letting $\beta_0 = \beta$ and $\beta_{n+1} = 1 + \alpha - \alpha/\beta_n$ for $n \geq 0$, 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.23), we obtain

$$
\frac{h_\varepsilon(\beta) - h_\varepsilon(\alpha)}{\beta - \alpha} \geq \frac{h_\varepsilon(\beta_n) - h_\varepsilon(\alpha)}{\beta_n - \alpha} \to \frac{h_\varepsilon(1) - h_\varepsilon(\alpha)}{1 - \alpha} = \frac{1}{1 - \alpha} \text{ as } n \to \infty
$$

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

$$
\min_{h \in H_{\xi_0}} h(\beta^*) > \frac{(\beta^* - \alpha)_+}{1 - \alpha}. \tag{EC.24}
$$

Meanwhile, by $h_\varepsilon(\beta) = 0$ for $\beta < \alpha$, we have $\min_{h \in H_\xi} h(\beta) = 0$ for any $\beta < \alpha$ and any $\xi \in \Xi$, and thus, $\min_{h \in H_{\xi_0}} h(\beta) = 0$ for $\beta < \alpha$. This implies that there exists $h_0 \in H_{\xi_0}$ such that $h_0(\beta) = 0$ for $\beta < \alpha$. By (EC.24), 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) \( \Rightarrow \) (c).

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

**Proof of Theorem EC.2.** Translation invariance, positive homogeneity and lower semicontinuity of \( \text{ES}_\alpha \), \( \alpha \in (0, 1) \), are well-known, and the property \( \preceq_2 \)-cEMA of \( \text{ES} \) follows from Theorem 1. Conversely, note that \( \preceq_2 \)-cEMA implies \( \preceq_2 \)-consistency, and hence, by Lemma EC.3, we have \( \rho = \text{ES}_\alpha \) 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 distribution \( F \). Since the probability space is atomless, 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

\[
U_n = \sum_{i=0}^{n-1} \frac{\alpha i}{n} 1_{\{\frac{\alpha i}{n} \leq U < \frac{\alpha (i+1)}{n}\}} + \sum_{i=0}^{n-1} \left( \alpha + \frac{(1 - \alpha) i}{n} \right) 1_{\{\alpha + \frac{(1 - \alpha) i}{n} \leq U < \alpha + \frac{(1 - \alpha) (i+1)}{n}\}}, \quad n \geq 1.
\]

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

**Remark EC.2.** The characterization results in Theorems EC.1 and EC.2 are obtained for spaces \( \mathcal{M}_p \), \( p \in [1, \infty) \), i.e., distributions with finite \( p \)th moment. On the space \( \mathcal{M}_\infty \) of compactly supported distributions, the situation is more delicate. In particular, for \( \alpha \in (0, 1) \) and \( \lambda \in (0, 1) \), we find that the mappings \( \lambda \text{VaR}_\alpha + (1 - \lambda) \text{VaR}_1 \) and \( \lambda \text{ES}_\alpha + (1 - \lambda) \text{VaR}_1 \) on \( \mathcal{M}_\infty \) satisfy the conditions in Theorems 2 and 3, 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.
E Proofs and generalizations for results in Section 5.2

E.1 A generalization of Theorem 4 and related results

The following theorem is a generalization of Theorem 4 to general space $M_p, p \in [0, \infty]$.

**Theorem EC.3.** A mapping $\rho : M_p \to \mathbb{R}, p \in [0, \infty]$ satisfies translation invariance, lower semicontinuity, $\rho(\delta_0) = 0$, and $\preceq_{1}$-EMA if and only if

$$\rho(F) = \sup_{\alpha \in [0,1)} \{\text{VaR}_\alpha(F) - h(\alpha)\}, \quad \text{(EC.25)}$$

for some increasing $h : [0, 1) \to [0, \infty]$ with $h(0^+) = 0$.

To show Theorem EC.3, we need the definition of acceptance set which will be also used in the proof of Theorem EC.4. It is well known that a $\preceq_{1}$-consistent and translation invariant risk measure, which is also called the monetary risk measure, can be characterized by an acceptance set; see Föllmer and Schied (2016). For a risk measure $\rho : M \to \mathbb{R}$, its acceptance set is defined as $A_\rho = \{F \in M : \rho(F) \leq 0\}$. The following lemma characterizes a class of acceptance sets by the property $\preceq_{1}$-EMA.

**Lemma EC.4.** Let $\rho : M_p \to \mathbb{R}, p \in [0, \infty]$, be a risk measure satisfying translation invariance. $\rho$ satisfies $\preceq_{1}$-EMA if and only if it satisfies $\preceq_{1}$-consistency, and $\bigvee_1 F \in A_\rho$ for any $F \subseteq A_\rho$ that is bounded from above.

**Proof.** The necessity is trivial. For sufficiency, let $F \subseteq M_p$ be bounded from above. The $\preceq_{1}$-consistency implies $\rho(\bigvee_1 F) \geq \sup_{F \in \mathcal{F}} \rho(F)$. On the other hand, denote by $\eta = \sup_{F \in \mathcal{F}} \rho(F) < \infty$, and define $\tilde{F} = \{\tilde{F} : \tilde{F}(\cdot) = F(\cdot + \eta), F \in \mathcal{F}\}$. It follows from the translation invariance of $\rho$ that $\tilde{F} \subseteq A_\rho$. Since $\inf_{F \in \tilde{F}} F(\cdot) = \inf_{F \in \mathcal{F}} F(\cdot + \eta)$, it follows from Proposition EC.1 that $\bigvee_1 \tilde{F}(\cdot) = \bigvee_1 F(\cdot + \eta)$. Hence, we have

$$\rho\left(\bigvee_1 F\right) = \rho\left(\bigvee_1 \tilde{F}\right) + \eta \leq \eta = \sup_{F \in \mathcal{F}} \rho(F),$$

where the first equality follows from translation invariance of $\rho$, and the inequality is due to $\bigvee_1 \tilde{F} \subseteq A_\rho$. Hence, we complete the proof.

**Proof of Theorem EC.3.** First consider sufficiency. It is straightforward to check that $\rho$ defined by (EC.25) satisfies $\preceq_{1}$-consistency, translation invariance and $\rho(\delta_0) = 0$. To see the lower semicontinuity of $\rho$, note that $\lim_{\alpha \to 0^+} \text{VaR}_\alpha(F) = \text{VaR}_0(F)$ and $h(0^+) = h(0)$, which implies
\( \rho(F) = \sup_{\alpha \in (0, 1)} \{ \text{VaR}_\alpha(F) - h(\alpha) \} \). It then follows from the lower semicontinuity of \( \text{VaR}_\alpha \), \( \alpha \in (0, 1) \), that \( \rho \) satisfies the lower semicontinuity. By the \( \preceq_1 \)-EMA of \( \text{VaR}_\alpha \), we have that \( \rho \) defined by (EC.25) satisfies \( \preceq_1 \)-EMA, which completes the proof of the sufficiency.

To show the necessity, we first verify that \( \rho \) has the form (EC.25) on \( \mathcal{M}_\infty \). Let \( \mathcal{A}_\rho \) be its acceptance set and define \( h(\alpha) := \sup_{G \in \mathcal{A}_\rho} \text{VaR}_\alpha(G), \alpha \in [0, 1) \). By \( \preceq_1 \)-consistency and translation invariance, to show (EC.25), it suffices to show \( F \in \mathcal{A}_\rho \) if and only if \( \text{VaR}_\alpha(F) \leq h(\alpha), \alpha \in [0, 1) \).

By definition of \( h \), it is straightforward to check that \( F \in \mathcal{A}_\rho \) implies \( \text{VaR}_\alpha(F) \leq h(\alpha), \alpha \in [0, 1) \).

For \( F \in \mathcal{M}_\infty \), if \( \text{VaR}_\alpha(F) \leq h(\alpha), \alpha \in [0, 1) \), then define \( \mathcal{F} = \{ F_X \wedge \text{VaR}_1(F) : F_X \in \mathcal{A}_\rho \} \). We have \( \mathcal{F} \subseteq \mathcal{A}_\rho \) by \( \preceq_1 \)-consistency, and \( \mathcal{F} \) is bounded from above as the constant \( \text{VaR}_1(F) \) serves an upper bound with respect to \( \preceq_1 \). By Lemma EC.4, we have \( \bigvee_1 \mathcal{F} \in \mathcal{A}_\rho \) and for \( \alpha \in [0, 1) \),

\[
\text{VaR}_\alpha(F) \leq \min \{ \text{VaR}_1(F), h(\alpha) \} = \sup_{G \in \mathcal{A}_\rho} \{ \text{VaR}_\alpha(G) \wedge \text{VaR}_1(F) \} = \sup_{G \in \mathcal{F}} \text{VaR}_\alpha(G) = \text{VaR}_\alpha \left( \bigvee_1 \mathcal{F} \right),
\]

that is, \( F \preceq_1 \bigvee_1 \mathcal{F} \in \mathcal{A}_\rho \). We thus have \( F \in \mathcal{A}_\rho \). Therefore, we have (EC.25) holds for \( h(\alpha) = \sup_{G \in \mathcal{A}_\rho} \text{VaR}_\alpha(G) \). Hence, \( \rho \) has the form (EC.25) on \( \mathcal{M}_\infty \).

By Lemma EC.1, \( \rho \) also has the form (EC.25) on \( \mathcal{M}_{bb} \). Let now \( F \in \mathcal{M}_\rho \setminus \mathcal{M}_{bb} \). We aim to show that \( \rho(F) \) has the representation (EC.25). To see this, denote by \( \zeta = \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(F) - h(\alpha) \} \) and \( H = \bigvee_1 \{ F, \delta_{\zeta - 1} \} \). Here we have \( \zeta \in \mathbb{R} \) since there exists \( G \in \mathcal{M}_{bb} \) such that \( F \preceq_1 G \) and thus \( \zeta = \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(G) - h(\alpha) \} = \rho(G) < \infty \), and by \( h(0+) = 0 \), there exists \( \alpha > 0 \) such that \( h(\alpha) \in \mathbb{R} \) and thus, \( \zeta > -\infty \). Note that \( H \in \mathcal{M}_{bb} \) since \( \delta_{\zeta - 1} \) is a degenerated distribution function. Thus, we have

\[
\rho(H) = \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(H) - h(\alpha) \} = \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(F) \lor \text{VaR}_\alpha(\delta_{\zeta - 1}) - h(\alpha) \} = \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(F) - h(\alpha) \} \lor \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(\delta_{\zeta - 1}) - h(\alpha) \} = \zeta \lor (\zeta - 1) = \zeta,
\]

where the second equality follows from the \( \preceq_1 \)-EMA of \( \text{VaR} \). Finally, by the property \( \preceq_1 \)-EMA of \( \rho \), we obtain

\[
\rho(F) \lor (\zeta - 1) = \rho(F) \lor \rho(\delta_{\zeta - 1}) = \rho(H) = \zeta,
\]

which implies \( \rho(F) = \zeta = \sup_{\alpha \in [0, 1)} \{ \text{VaR}_\alpha(F) - h(\alpha) \} \), and hence \( \rho \) has the form (EC.25) on \( \mathcal{M}_\rho \).
Note that $h$ is increasing, and $\rho(\delta_0) = -\inf_{\alpha \in [0,1]} h(\alpha) = 0$ implies $h(0) = 0$. It remains to see that $h(0+) = h(0)$. Assume by contradiction that $h(0+) > h(0)$. Let $F_n = (1/n)\delta_n + (1 - 1/n)\delta_0$ for $n \geq 1$. It is easy to see that $F_n \xrightarrow{d} \delta_0$. We can calculate that $\rho(F_n) = (-n) \vee (-h((1/n)+))$ for $n \geq 1$. Thus, $\liminf_{n \to \infty} \rho(F_n) = -h(0+) < -h(0) = \rho(\delta_0)$, which contradicts with the lower semicontinuity of $\rho$. We thus complete the proof. \qed

Remark EC.3. In Theorem EC.3, we assumed that $\rho$ is real-valued; however $\rho$ in (EC.25) does not always define a real-valued mapping. Specifically, we have

(i) On $\mathcal{M}_\infty$, $\rho$ in (EC.25) is always real-valued. This is straightforward to see.

(ii) On $\mathcal{M}_0$, $\rho$ in (EC.25) is real-valued if and only if $h(\alpha) = \infty$ for some $\alpha \in (0,1)$. To show the “only if” statement, suppose now $h(\alpha) < \infty$ for all $\alpha \in (0,1)$, and let $F \in \mathcal{M}_0$ such that $\text{VaR}_\alpha(F_0) = h(\alpha) + 1/(1-\alpha)$. Obviously, we have $\limsup_{\alpha \to 1} \{\text{VaR}_\alpha(F) - h(\alpha)\} = \infty$ which implies $\sup_{\alpha \in (0,1)} \{\text{VaR}_\alpha(F) - h(\alpha)\} = \infty$. To show the if part, assume that $h(\alpha_0) = \infty$ for some $\alpha_0 \in (0,1)$. Then for any $F \in \mathcal{M}_0$, we have $\sup_{\alpha \in [0,1)} \{\text{VaR}_\alpha(F) - h(\alpha)\} \leq \text{VaR}_{\alpha_0}(F) < \infty$. Therefore, we have the statement holds.

(iii) On $\mathcal{M}_p$, $p \in (0,\infty)$, a necessary condition for that $\rho$ in (EC.25) is real-valued is $\int_0^1 (h(\alpha))^p d\alpha = \infty$. To see this, suppose that $\int_0^1 (h(\alpha))^p d\alpha < \infty$, and let $F_0$ be a distribution such that $\text{VaR}_\alpha(F_0) = h(\alpha) + (1-\alpha)^{-1/(2p)}$. We have $F_0 \in \mathcal{M}_p$ and $\limsup_{\alpha \to 1} \{\text{VaR}_\alpha(F_0) - h(\alpha)\} = \infty$ which implies $\sup_{\alpha \in (0,1)} \{\text{VaR}_\alpha(F_0) - h(\alpha)\} = \infty$. Hence, $\int_0^1 (h(\alpha))^p d\alpha = \infty$ is a necessary condition.

The following corollary is a generalized result of Corollary 1 for the space $\mathcal{M}_p$, $p \in [0,\infty]$. 

Corollary EC.1. Let $\rho : \mathcal{M} \to \mathbb{R}$ be a mapping satisfying translation invariance, positive homogeneity and lower semicontinuity.

(a) If $\mathcal{M} = \mathcal{M}_\infty$, then $\rho$ satisfies $\preceq_1$-EMA if and only if $\rho = \text{VaR}_\alpha$ for some $\alpha \in (0,1]$.

(b) If $\mathcal{M} = \mathcal{M}_p$, $p \in [0,\infty)$, then $\rho$ satisfies $\preceq_1$-EMA if and only if $\rho = \text{VaR}_\alpha$ for some $\alpha \in (0,1)$.

Proof. It can be immediately verified by applying Theorem EC.3 and Proposition 4.6 of Bignozzi et al. (2020). \qed

E.2 A generalization of Theorem 5 and related results

In this appendix, we aim to establish a generalization of Theorem 5 in Section 5.2 to the domain $\mathcal{M}_p$, $p \in [1,\infty]$. By Proposition EC.1, the ordered sets $(\mathcal{M}_p, \preceq_2)$, $p \in (1,\infty]$, are not
complete lattice which means the supremum may not exist even if the uncertainty set is bounded from above. This observation suggests that it is no longer natural to define $\preceq_2$-EMA on $\mathcal{M}_p$, $p \in (1, \infty]$, via equation (14). In this case, noting that the ordered sets are all lattice, we propose a definition of equivalence in finite model aggregation as follows.

$\preceq_2$-fEMA: For a lattice $(\mathcal{M}, \preceq)$, a mapping $\rho : \mathcal{M} \to \mathbb{R}$ is said to be equivalent in finite model aggregation if (14), i.e., $\rho(\bigvee \mathcal{F}) = \sup_{F \in \mathcal{F}} \rho(F)$, holds for any $\mathcal{F} \in \mathcal{M}$ which contains finitely many distributions.

The following theorem identifies the class of risk measures that are characterized by $\preceq_2$-fEMA on $\mathcal{M}_p$, $p \in [1, \infty]$.

**Theorem EC.4.** A mapping $\rho : \mathcal{M}_p \to \mathbb{R}$, $p \in [1, \infty]$, satisfies $\rho(\delta_0) = 0$, translation invariance, and $\preceq_2$-fEMA if and only if

$$\rho(F) = \sup_{\alpha \in [0,1]} \{E\alpha(F) - g(\alpha)\}, \quad (\text{EC.26})$$

for some increasing $g : [0,1) \to [0, \infty]$ with $g(0) = 0$ such that $h(\alpha) := (1-\alpha)g(\alpha)$ is concave.

In Theorem EC.4, it is assumed that $\rho$ is real-valued. Similarly to Theorem EC.3 (see Remark EC.3), there are some conditions for (EC.26) to define a real-valued mapping when $p \in [1, \infty)$. A sufficient condition for $\rho$ in (EC.26) to be real-valued is that $h(1-) > 0$, which is also a necessary condition for $\rho$ in (EC.26) to be real-valued on $\mathcal{M}_1$.

Below, we will first prove Theorem EC.4, and then apply the result in Theorem EC.4 to prove Theorem 5. In order to prove Theorem EC.4, we first show the following proposition that translation invariant risk measure satisfying $\preceq_2$-fEMA is automatically convex\(^\text{14}\).

**Proposition EC.2.** Let $(\mathcal{M}, \preceq_2)$ be a lattice. Every translation invariant mapping $\rho : \mathcal{M} \to \mathbb{R}$ satisfying $\preceq_2$-fEMA is convex.

**Proof.** For $F_X, F_Y \in \mathcal{M}$ and $\lambda \in (0, 1)$, since $\pi_{F_{\lambda X + (1-\lambda)Y}}(t) \leq \lambda \pi_{F_X}(t) + (1-\lambda)\pi_{F_Y}(t) \leq \pi_{F_X}(t) \lor \pi_{F_Y}(t)$ for all $t \in \mathbb{R}$, we have $F_{\lambda X + (1-\lambda)Y} \preceq_2 \bigvee_{2} \{F_X, F_Y\}$ for all $\lambda \in (0, 1)$. Denote by $x = \rho(F_X)$ and $y = \rho(F_Y)$. By translation invariance and $\preceq_2$-fEMA of $\rho$, we have

$$\rho(F_{\lambda X + (1-\lambda)Y}) - \lambda x - (1-\lambda)y = \rho(F_{\lambda X + (1-\lambda)Y} - \lambda x - (1-\lambda)y) \leq \rho \left( \bigvee_{2} \{F_{X-x}, F_{Y-y}\} \right) = \rho(F_{X-x}) \lor \rho(F_{Y-y}) = 0,$$

\(^{14}\)Here, convexity of a risk measure means that it is a convex functional when treated as a mapping from a space of random variables $\mathcal{X}$ (instead of $\mathcal{M}$) to $\mathbb{R}$.
where the inequality holds because \( \preceq_2 \)-fEMA implies \( \preceq_2 \)-consistency. Hence, we complete the proof.

Moreover, we need some other preliminaries as follows.

1. Integrated quantile function of the supremum with respect to \( \preceq_2 \): Similar to Proposition EC.1, we give the explicit form of \( \sup_2 F \) in terms of quantile function. Before showing the result, recall the definition of integrated quantile function in (EC.2) that for \( F \in M_1 \), \( E_F(\alpha) = (1 - \alpha)E\alpha(F) = \int_0^1 F^{-1}(s)ds \) for \( \alpha \in [0, 1] \). The set of all integrated quantile functions contains all real-valued functions \( g \) that are concave, have no jump at zero and \( g(1) = 0 \). Moreover, the quantile function \( F^{-1} \) can be uniquely derived by \( F^{-1}(\alpha) = -(E_F)'_-(\alpha) \) for \( \alpha \in [0, 1) \), where \( (E_F)'_-(\alpha) \) is the left derivative of \( E_F \). The concave envelope of a function \( f \) is defined as

\[
 f^*(x) = \inf \{g(x) : g \text{ is concave and } g(y) \geq f(y) \text{ for all } y \in \mathbb{R}\},
\]

yielding the smallest concave function larger than a given one.

**Proposition EC.3.** For \( F \subseteq M_1 \) which is bounded from above, we have \( E\sup_2 F = (\sup_{F \in F} E_F)^* \) on \((0, 1)\) and \( (\sup_2 F)^{-1} = -((\sup_{F \in F} E_F)^*)'_- \) on \((0, 1)\).

**Proof.** Note that \( F \preceq_2 G \) if and only if \( E_F(\alpha) \leq E_G(\alpha) \) for all \( \alpha \in (0, 1) \), and \( \sup_2 F \in M_1 \) exists by Proposition EC.1 (b). We have \( E\sup_2 F = (\sup_{F \in F} E_F)^* \) on \((0, 1)\). The left quantile function of \( \sup_2 F \) follows from the fact that \( F^{-1} = -(E_F)'_-(\alpha) \) for all \( F \in M \). \qed

2. The acceptance set for the risk measure that satisfies \( \preceq_2 \)-(f)EMA: Recall that a \( \preceq_1 \)-consistent and translation invariant risk measure \( \rho : M \to \mathbb{R} \) can be characterized by an acceptance set, i.e., \( A_\rho = \{ F \in M : \rho(F) \leq 0 \} \). The following lemma characterizes a class of acceptance sets by the property \( \preceq_2 \)-(f)EMA.

**Lemma EC.5.** Let \( \rho : M \to \mathbb{R} \) be a risk measure satisfying translation invariance.

(a) If \( M = M_p \), \( p \in (1, \infty] \), then \( \rho \) satisfies \( \preceq_2 \)-fEMA if and only if \( \rho \) satisfies \( \preceq_2 \)-consistency, and \( \sup_2 \{F, G\} \in A_\rho \) for any \( F, G \in A_\rho \).

(b) If \( M = M_1 \), then \( \rho \) satisfies \( \preceq_2 \)-EMA if and only if \( \rho \) satisfies \( \preceq_2 \)-consistency, and \( \sup_2 F \in A_\rho \) for any \( F \subseteq A_\rho \) that is bounded from above.

**Proof.** We only prove (b) as the proof of (a) is similar. The necessity is trivial. To see sufficiency, for \( F \subseteq M_1 \) which is bounded from above, the property \( \preceq_2 \)-consistency implies \( \rho(\sup_2 F) \geq \rho(F) \). On the other hand, denote by \( \eta = \sup_{F \in F} \rho(F) < \infty \), and define \( \tilde{F} = \{ \tilde{F} : \rho(\tilde{F}) < \eta \} \).
\( \tilde{F}(\cdot) = F(\cdot + \eta), \ F \in \mathcal{F} \). It follows from the translation invariance of \( \rho \) that \( \tilde{\mathcal{F}} \subseteq \mathcal{A}_\rho \). Since \( \sup_{F \in \tilde{\mathcal{F}}} \pi_F(\cdot) = \sup_{F \in \mathcal{F}} \pi_F(\cdot + \eta) \), it follows from Proposition EC.1 that \( \sqrt{2} \tilde{F}(\cdot) = \sqrt{2} F(\cdot + \eta) \). Hence, we have

\[
\rho \left( \sqrt{2} \mathcal{F} \right) = \rho \left( \sqrt{2} \tilde{\mathcal{F}} \right) + \eta \leq \eta = \sup_{F \in \mathcal{F}} \rho(F),
\]

where the first equality follows from translation invariance of \( \rho \), and the inequality is due to \( \sqrt{2} \tilde{\mathcal{F}} \subseteq \mathcal{A}_\rho \). Hence, we complete the proof.

3. **Kusuoka representation:** It is worth notice that law-invariant convex risk measure admits a Kusuoka representation; see Kusuoka (2001). For a translation invariant risk measure \( \rho : \mathcal{M}_\infty \to \mathbb{R} \) that satisfies \( \preceq_{2\text{-fEMA}} \), we know that \( \rho \) is a convex risk measure (see Proposition EC.2), and hence, define its Kusuoka representation as

\[
\rho(F) = \sup_{\mu \in \mathcal{M}([0,1])} \left\{ \int_{[0,1]} \text{ES}_\alpha(F) \mu(d\alpha) - R_\rho(\mu) \right\},
\]

where \( \mathcal{M}([0,1]) \) is the space of all probability measures on \( ([0,1], \mathcal{B}([0,1])) \) and \( R_\rho(\mu) \) is the penalty function which can be represented as

\[
R_\rho(\mu) = \sup_{G \in \mathcal{A}_\rho} \int_{[0,1]} \text{ES}_\alpha(G) \mu(d\alpha).
\]

**Proof of Theorem EC.4.** To see sufficiency, note that \( \rho \) defined by (EC.26) satisfies \( \preceq_{2\text{-fEMA}} \)-consistency and \( \rho(\delta_0) = 0 \). The acceptance set of \( \rho \) is

\[
\mathcal{A}_\rho = \{ F \in \mathcal{M}_\infty : E_F(\alpha) - h(\alpha) \leq 0 \ \text{for all} \ \alpha \in [0,1] \},
\]

where \( E_F \) is the integrated quantile function defined in (EC.2). For \( F, G \in \mathcal{A}_\rho \), it follows from Proposition EC.3 that \( E_{\sqrt{2}\{F,G\}} = (E_F \lor E_G)^* \). Since \( h(\alpha) \) is concave on \( [0,1] \), we have \( (E_F \lor E_G)^*(\alpha) \leq h(\alpha) \) for all \( \alpha \in [0,1] \), which implies \( \sqrt{2}\{F,G\} \in \mathcal{A}_\rho \). By Lemma EC.5, we have \( \rho \) satisfies \( \preceq_{2\text{-fEMA}} \), and thus we complete the proof of sufficiency.

To see necessity, by Proposition EC.2, we know that \( \rho \) is a convex risk measure so that \( \rho \) admits a Kusuoka representation. Denote by \( R_\rho \) the penalty function of \( \rho \), and define

\[
g(\alpha) := R_\rho(\delta_\alpha) = \sup_{F \in \mathcal{A}_\rho} \text{ES}_\alpha(F), \ \alpha \in [0,1].
\]
One can check that $g$ is increasing on $[0, 1)$. In the following, we aim to show that $g$ is the candidate function in (EC.26). First, we will verify that $h(\alpha) := (1 - \alpha)g(\alpha) = \sup_{F \in \mathcal{A}_\rho} E_F(\alpha)$ is concave on $[0, 1)$. By contradiction, there exist $0 \leq \alpha_1 < \alpha_2 < 1$ such that $h(\frac{\alpha_1 + \alpha_2}{2}) < \frac{h(\alpha_1)}{2} + \frac{h(\alpha_2)}{2}$.

Hence, there exist $F_1, F_2 \in \mathcal{A}_\rho$ such that $\frac{E_{F_1}(\alpha_1)}{2} + \frac{E_{F_2}(\alpha_2)}{2} > \sup_{F \in \mathcal{A}_\rho} E_F(\frac{\alpha_1 + \alpha_2}{2})$. Denote by $F_0 = \bigvee \{F_1, F_2\}$. By Lemma EC.5, we know that $F_0 \in \mathcal{A}_\rho$. Therefore,

$$\frac{E_{F_1}(\alpha_1)}{2} + \frac{E_{F_2}(\alpha_2)}{2} > E_{F_0}\left(\frac{\alpha_1 + \alpha_2}{2}\right) \geq \frac{E_{F_1}(\alpha_1)}{2} + \frac{E_{F_2}(\alpha_2)}{2},$$

where the second inequality follows from the concavity of $E_{F_0}$. This yields a contradiction to that $E_{F_1}(\alpha) \vee E_{F_2}(\alpha) \leq E_{F_0}(\alpha)$ for all $\alpha \in [0, 1)$. Hence, we have $h$ is concave on $[0, 1)$. It follows from Lemma EC.5 that $F, G \in \mathcal{A}_\rho$ implies $\bigvee \{F, G\} \in \mathcal{A}_\rho$. Hence, by Theorem A.33 of Föllmer and Schied (2016), there exists an increasing sequence $E_{F_n}(\alpha) \uparrow h(\alpha)$ for all $\alpha \in \mathcal{Q}([0, 1))$ where $\{F_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_\rho$. Since $h$ and $E_{F_n}$ for $n \geq 1$ are all concave functions, we have $E_{F_n}(\alpha) \uparrow h(\alpha)$ for all $\alpha \in [0, 1)$. On one hand, by monotone convergence theorem, we have

$$R_\rho(\mu) = \sup_{F \in \mathcal{A}_\rho} \int_{[0, 1)} ES_\alpha(F)\mu(d\alpha) \geq \lim_{n \to \infty} \int_{[0, 1)} ES_\alpha(F_n)\mu(d\alpha) = \int_{[0, 1)} g(\alpha)\mu(d\alpha).$$

On the other hand, we have

$$R_\rho(\mu) = \sup_{F \in \mathcal{A}_\rho} \int_{[0, 1)} ES_\alpha(F)\mu(d\alpha) \leq \int_{[0, 1)} \sup_{F \in \mathcal{A}_\rho} ES_\alpha(G)\mu(d\alpha) = \int_{[0, 1)} g(\alpha)\mu(d\alpha).$$

Hence, we obtain $R_\rho(\mu) = \int_{[0, 1)} g(\alpha)\mu(d\alpha)$. Substituting the penalty function into the Kusuoka representation, we have for $F \in \mathcal{M}_\infty$,

$$\rho(F) = \sup_{\mu \in \mathcal{M}([0, 1))} \left\{ \int_{[0, 1)} ES_\alpha(F)\mu(d\alpha) - R_\rho(\mu) \right\}$$

$$= \sup_{\mu \in \mathcal{M}([0, 1))} \left\{ \int_{[0, 1)} ES_\alpha(F)\mu(d\alpha) - \int_{[0, 1)} g(\alpha)\mu(d\alpha) \right\}$$

$$= \sup_{\mu \in \mathcal{M}([0, 1))} \left\{ \int_{[0, 1)} (ES_\alpha(F) - g(\alpha))\mu(d\alpha) \right\}$$

$$= \sup_{\alpha \in [0, 1)} \{ES_\alpha(F) - g(\alpha)\}.$$

Since $\rho(\delta_0) = 0$, we have $g(0) = \inf_{\alpha \in [0, 1]} g(\alpha) = 0$. Hence, we have that $\rho$ have formula (EC.26) on $\mathcal{M}_\infty$. By Proposition EC.2, we know that $\rho$ is convex, and it follows from Theorem 2.2 of Filipović and Svindland (2012) that $\rho$ can be uniquely extended from $\mathcal{M}_\infty$ to $\mathcal{M}_p$, $p \in [1, \infty)$. This completes
the proof.

Proof of Theorem 5. To see sufficiency, note that $\rho$ with form (17) satisfies translation invariance, $\preceq_2$-consistency and $\rho(\delta_0) = 0$. The acceptance set of $\rho$ is

$$A_\rho = \{ F \in \mathcal{M}_1 : E_F(\alpha) - h(\alpha) \leq 0 \text{ for all } \alpha \in [0, 1) \}.$$ 

By Lemma EC.5, it suffices to show that $\bigvee_2 F \in A_\rho$ for any $F \subseteq A_\rho$ such that $F$ is bounded from above. To see this, let now $F \subseteq A_\rho$ which is bounded from above. Since $h(\alpha)$ is concave and continuous at zero, it follows from Proposition EC.3 that $E_{\bigvee_2 F}(\alpha) = (\sup_{F \in F} E_F(\alpha))^*(\alpha) \leq h(\alpha)$ for all $\alpha \in [0, 1)$, which implies $\bigvee_2 F \in A_\rho$. This completes the proof of sufficiency.

To see necessity, note that $\preceq_2$-EMA is stronger than $\preceq_2$-fEMA. Hence, applying Theorem EC.4, we know that $\rho$ has the form (17), and it suffices to show $g(0^+) = 0$ and $h(1^-) > 0$.

Assume by contradiction that $g(0^+) > 0$. Let $\{F_n\}_{n \in \mathbb{N}}$ be a sequence of distributions such that $F_n = (1/n)\delta_{-n} + ((n - 1)/n)\delta_0$. One can verify that $\bigvee_2 \{F_n\}_{n \in \mathbb{N}} = \delta_0$. Consider,

$$\rho(F_n) = \sup_{\alpha \in [0, 1)} \{ ES_\alpha(F_n) - g(\alpha) \}$$

$$= \left( \sup_{\alpha \in [0, 1/n]} \left\{ \frac{n\alpha - 1}{1 - \alpha} - g(\alpha) \right\} \right) \lor \left( \frac{1}{n} \right), \quad n \geq 1.$$

Note that the function $\alpha \mapsto (n\alpha - 1)/(1 - \alpha)$ is nonpositive for $\alpha \in [0, 1/n]$, and takes the value $-1$ if $\alpha = 0$. Therefore, we obtain $\rho(F_n) \leq (1) \lor (-g(0^+)) < 0$ for all $n$. Hence, we have

$$\rho \left( \bigvee_2 \{F_n\}_{n \in \mathbb{N}} \right) = \rho(\delta_0) = 0 > (1) \lor (-g(0^+)) \geq \sup_{n \geq 1} \rho(F_n),$$

which means that $\rho$ does not satisfy $\preceq_2$-EMA, a contradiction. Hence, we have $g(0^+) = 0$.

Suppose by contradiction that $h(1^-) = 0$. Let $F \in \mathcal{M}_1$ such that $E_F(\alpha) = h(\alpha) + \sqrt{1 - \alpha}$ for $\alpha \in (0, 1)$. We obtain

$$\rho(F) = \sup_{\alpha \in [0, 1]} \frac{1}{1 - \alpha} (E_F(\alpha) - h(\alpha)) \geq \sup_{\alpha \in (0, 1)} \frac{\sqrt{1 - \alpha}}{1 - \alpha} = \infty.$$ 

This yields a contradiction. Hence, we have $h(1^-) > 0$ and complete the proof.  

\[57\]
F Proofs for results in Section 6 and omitted figures

F.1 Proofs of Theorems 6 and 7

**Proof of Theorem 6.** 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 \( \text{sup}_{F \in \mathcal{F}_{1, \varepsilon}(F_0)} \pi_F(x) = \pi_{F_0}(x) + \varepsilon \) for all \( x \in \mathbb{R} \). There does not exist \( G \in \mathcal{M}_1 \) such that \( \pi_G = \text{sup}_{F \in \mathcal{F}_{1, \varepsilon}(F_0)} \pi_F \) since \( \lim_{x \to \infty} \text{sup}_{F \in \mathcal{F}_{1, \varepsilon}(F_0)} \pi_F(x) = \varepsilon > 0 \). Hence, the set \( \mathcal{F}_{1, \varepsilon}(F_0) \) is not bounded from above with respect to \( \leq_2 \). For \( p > 1 \), since the Wasserstein ball is convex, it follows from Theorem 1 that \( \text{sup}_{F \in \mathcal{F}_{p, \varepsilon}(F_0)} \text{ES}_\alpha(F) = \text{ES}_\alpha(F_{p, \varepsilon}^2(F_0)) \). By Proposition 4 (ii) of Liu et al. (2022), we have \( \text{sup}_{F \in \mathcal{F}_{p, \varepsilon}(F_0)} \text{ES}_\alpha(F) = \text{ES}_\alpha(F_0) + (1 - \alpha)^{-1/p} \varepsilon \) for \( \alpha \in (0, 1) \). Therefore, one can obtain

\[
\int_0^1 \text{VaR}_s \left( F_{p, \varepsilon}^2(F_0) \right) ds = \int_0^1 \text{VaR}_s(F_0) ds + (1 - \alpha)^{1 - \frac{1}{p}} \varepsilon, \quad \alpha \in (0, 1).
\]

Take the derivative on the left and right sides of the above formula for \( \alpha \), we have

\[
\text{VaR}_\alpha \left( F_{p, \varepsilon}^2(F_0) \right) = \text{VaR}_\alpha(F_0) + \left( 1 - \frac{1}{p} \right) (1 - \alpha)^{-\frac{1}{p}} \varepsilon.
\]

Hence, we complete the proof.

**Proof of Theorem 7.** For two random vectors \( X \) and \( Y \) of the same dimension, define \( \mathcal{L}_p \) as

\[
\mathcal{L}_p(X, Y)^p = \mathbb{E} [\|X - Y\|^p].
\]

For any \( F \in \mathcal{F}_{w, p, \varepsilon}(F_X) \), by definition, there exists \( Z \) with \( F_Z \in \mathcal{F}_{p, \varepsilon}^d(F_X) \) and \( F = F_{w \top Z} \). We can verify that

\[
W_p(F_{w \top X}, F_{w \top Z}) = \inf_{X \in \mathcal{F}_{w \top X}, Z \in \mathcal{F}_{w \top Z}} (\mathbb{E}[\|X - Z\|_p]^p)^{1/p} \leq \inf_{\|w\|_\infty \mathcal{L}_p(X', Z')} \|w\|_\infty \mathcal{L}_p(X', Z') = \|w\|_\infty W_p(F_X, F_Z) \leq \|w\|_\infty \varepsilon,
\]

where the infima are taken over \((X, Z)\) or \((X', Z')\), and the first inequality follows from the Hölder inequality. Hence, \( \mathcal{F}_{w, p, \varepsilon}(F_X) \subseteq \mathcal{F}_{p, \|w\|_\infty \varepsilon}(F_{w \top X}) \).

We next show the opposite direction of the set inclusion. For any \( F \in \mathcal{F}_{p, \|w\|_\infty \varepsilon}(F_{w \top X}) \),
Figure EC.1: The supremum of $F_{2,0.1}(F_0)$ with $F_0 \sim N(0,1)$.

since the set $\{Y : (E[|Y - w^\top X|^p])^{1/p} \leq \|w\|_* \varepsilon\}$ is closed, there exists $Z \sim F$ such that $(E[|Z - w^\top X|^p])^{1/p} \leq \|w\|_* \varepsilon$. Denote by $Y = Z - w^\top X$, and we have $E[|Y|^p] \leq \|w\|_*^p \varepsilon^p$. By the definition of the dual norm, there exists $x_0 \in \mathbb{R}^d$ such that $\|w\|_* = x_0^\top w$ and $\|x_0\| = 1$. Define $Z = X + (x_0 Y)/\|w\|_*$, and we have

$$\left(W_p^d(F_X, F_Z)\right)^p \leq E[\|Z - X|^p] = E \left[ \left( \frac{x_0}{\|w\|_*} Y \right)^p \right] = E \left[ \frac{\|x_0\|^p}{\|w\|_*^p} |Y|^p \right] = \frac{E[|Y|^p]}{\|w\|_*^p} \leq \varepsilon^p.$$  

Hence, $F_Z \in F_{p,\varepsilon}^d(F_X)$ which implies $F_{w^\top Z} \in F_{w,p,\varepsilon}(F_X)$. Note that $w^\top Z = w^\top X + Y = Z$, we have $F \in F_{w,p,\varepsilon}(F_X)$, and this concludes that $F_{w,p,\varepsilon}(F_X) \supseteq F_{p,\|w\|_* \varepsilon}(F_{w^\top X})$.

F.2 Omitted figures from Section 6

We present a few figures omitted from Section 6. Figures EC.1 and EC.2 are related to Section 6.1 for the Wasserstein uncertainty set $F_{k,\varepsilon}(F_0)$ with $k = 2$, $\varepsilon = 0.1$, where the baseline distribution $F_0$ is the standard normal distribution function. Figure EC.1 shows the left quantile functions of the supremum. In Figure EC.2, we obtain the WR and MA robust risk values and the risk measure is chosen as ES$_\alpha$ or PD$_k$.

Figure EC.3 shows the curves of robust risk evaluation via the WR and MA approaches for the mean-variance uncertainty set $F_{0,1}$ in Section 6.3. The risk measure is chosen as ES$_\alpha$, RVaR$_{\alpha,\beta}$, PD$_k$ or ex$_\alpha$.  

59
Figure EC.2: The WR and MA approaches under $F_{2,0.1}(F_0)$ with $F_0 \sim N(0,1)$.

G Supplementary numerical results in Section 7

G.1 Portfolio selection under mean-variance uncertainty

We follow the portfolio selection setting discussed in Section 7.2 to assume that only the mean and the covariance matrix are available to the investor. This appendix complements the study in Section 7.2 with Wasserstein uncertainty.

For a given portfolio weight $w$, the uncertainty set is $F_{w^\top \mu, \sqrt{w^\top \Sigma}w}$, where $\mu$ is the mean vector and $\Sigma$ is the covariance matrix of losses from the stocks as reported in or computed from Table 2. 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

$$
\min_{w \in \Delta_5} : \rho_{\preceq_1}^{MA} \left( F_{w^\top \mu, \sqrt{w^\top \Sigma}w} \right) = w^\top \mu + \beta_k \sqrt{w^\top \Sigma w} \quad \text{s.t.} \quad w^\top \mu \leq -r_0,
$$

(EC.27)

and

$$
\min_{w \in \Delta_5} : \rho_{\preceq_2}^{MA} \left( F_{w^\top \mu, \sqrt{w^\top \Sigma}w} \right) = w^\top \mu + \gamma_k \sqrt{w^\top \Sigma w} \quad \text{s.t.} \quad w^\top \mu \leq -r_0,
$$

(EC.28)

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. Using results of Li (2018), the WR portfolio optimization problem is

$$
\min_{w \in \Delta_5} : \rho^{WR} \left( F_{w^\top \mu, \sqrt{w^\top \Sigma}w} \right) = w^\top \mu + \eta_k \sqrt{w^\top \Sigma w} \quad \text{s.t.} \quad w^\top \mu \leq -r_0,
$$

(EC.29)
where \( \eta_k = (k - 1)/\sqrt{2k - 1} \). Figure EC.4 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.

Similarly to Section 7.2, we choose 350 trading days for the initial training, and compute the optimal portfolio weights in each day with a rolling window. Figure EC.5 presents the three approaches under mean-variance uncertainty over the remaining 300 trading days with \( r_0 = 0.0015 \) and \( k = 2 \) (left) or \( k = 20 \) (right). In both cases, the MA approaches with \( \preceq 1 \) and \( \preceq 2 \) have very similar performance. In the case \( k = 2 \), the SAA approach outperforms the other approaches over most of the time period, with MA slightly better than WR after the first 150 trading days. In the
Figure EC.4: The optimized values of PD_k under mean-variance uncertainty. Left: r_0 = 0.0015 and k ∈ [1, 20]; Right: k = 10 and r_0 ∈ [0.0005, 0.0023].

Figure EC.5: Wealth evolution under mean-variance uncertainty with r_0 = 0.0015. 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.27), (EC.28) and (EC.29) also grow. If these weights are large enough, then the terms involving √w^TΣw in those problems become dominant in the optimization. For this reason, problems (EC.27), (EC.28) and (EC.29) are similar to the classical mean-variance optimization.
Figure EC.6: The optimized values of $PD_k$ under Wasserstein uncertainty. Left: $r_0 = 0.0015$, $k = 10$ and $\varepsilon \in [0, 0.03]$; Middle: $r_0 = 0.0015$, $\varepsilon = 0.01$ and $k \in [1, 20]$; Right: $k = 10$, $\varepsilon = 0.01$ and $r_0 \in [0.0005, 0.0023]$.

with a large weight on the variance.

G.2 Wasserstein uncertainty with a normal benchmark distribution

We follow the portfolio selection setting discussed in Section 7.2 under Wasserstein uncertainty with a fitted normal benchmark distribution. The considered optimization problems have the same form of (31) and (32) with the unit variance $t$-distribution replaced by the standard norm distribution. Figure EC.6 presents the robust risk values of the optimization problem with the SAA, WR and MA approaches for different values of $\varepsilon$, $r_0$ and $k$ using the whole-period data. In the left panel, we see that SSA may be the largest if $\varepsilon \leq 0.001$, because the empirical distribution of the data may be outside the Wasserstein uncertainty set if $\varepsilon$ is too small. As seen from Theorem 6, although the multivariate normal distribution of $X$ leads to a light-tailed benchmark distribution of $w^T X$, the robust model we use in the MA approach is heavy-tailed. Figure EC.7 reports the wealth process with a normal benchmark distribution, which shows a similar pattern to Figure 6 in Section 7.2.

H Omitted technical remarks from the main paper

Remark EC.5 is related to the technical conditions in Theorem 4 in Section 5.2. Remarks EC.6 and EC.7 are related to the technical conditions in Theorem 5 in Section 5.2. Remark EC.8 is related to the robust distributions in Theorem 6 in Section 6.

Remark EC.5. To guarantee that (16) defines a real-valued risk measure on $\mathcal{M}_1$, some conditions on $h : [0, 1] \rightarrow [0, \infty]$ need to be imposed. It is necessary that $\int_0^1 h(\alpha) d\alpha = \infty$, and it is sufficient if
Figure EC.7: Wealth evolution under Wasserstein uncertainty with $\varepsilon = 0.01$ and $r_0 = 0.0015$. Left: $k = 2$; Right: $k = 20$.

$h(\alpha) = \infty$ for some $\alpha \in (0, 1)$. If we allow $\rho$ to take $\infty$, that is, $\rho : \mathcal{M}_1 \to \mathbb{R} \cup \{\infty\}$, then the result in Theorem 4 also holds true.

Remark EC.6. Different from Theorems 2, 3 and 4, lower semicontinuity is not assumed in Theorem 5. This is because translation invariance and $\preceq_{2}$-EMA of $\rho$ imply that $\rho$ is a convex risk measure (see Proposition EC.2 in the appendix), which further implies that $\rho$ is $L^1$-continuous; see Kaina and Rüschendorf (2009), and $L^1$-continuity in place of lower semicontinuity is enough to complete the proof of Theorem 5. Note that $L^1$-continuity does not directly imply lower semicontinuity with respect to $\implies$.

Remark EC.7. As mentioned in Remark 4, $h(1-) > 0$ ensures that $\rho$ defined by (17) is real-valued on $\mathcal{M}_1$. If we allow $\rho$ to take $\infty$, that is, $\rho : \mathcal{M}_1 \to \mathbb{R} \cup \{\infty\}$, then the characterization in Theorem 5 remains true by removing the constraint $h(1-) > 0$, following from the same proof. The constraint $g(0+) = 0$ implies both $g(0) = 0$ and $g(0+)$ leading to two implications.

1. The first constraint $g(0) = 0$ is used only to ensure $\rho(\delta_0) = 0$. Indeed, if we instead assume $\rho(\delta_0) = a$ with some $a \in \mathbb{R}$, this constraint can be replaced by $g(0) = a$, and the statements in Theorem 5 remain valid.

2. The continuity condition $g(0+) = g(0)$ implies lower semicontinuity of $\rho$ in (17), and it is essential to $\preceq_{2}$-EMA. To see this, let $\{F_n\}_{n \in \mathbb{N}}$ be a sequence such that $F_n = (1/n)\delta_{-n} + ((n-1)/n)\delta_0$. One can easily verify that $\bigvee_2 \{F_n\}_{n \in \mathbb{N}} = \delta_0$. Therefore, if $g$ has a jump at zero, then
we have (see the proof of Theorem 5 in Appendix E for details)

\[
\rho \left( \bigvee_{n \in \mathbb{N}} \{ F_n \} \right) = \sup_{\alpha \in [0,1)} \{ \text{ES}_\alpha(\delta_0) - g(\alpha) \}
\]

\[
> \sup_{n \in \mathbb{N}} \sup_{\alpha \in [0,1)} \{ \text{ES}_\alpha(F_n) - g(\alpha) \} = \sup_{n \in \mathbb{N}} \rho(F_n),
\]

which implies that \( \rho \) does not satisfy \( \preceq_2 \)-EMA.

Remark EC.8. 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 6.

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

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

(iii) The left-hand side of equation (20) is increasing in \( p \). Hence, a larger value of \( p \) leads to a smaller distribution function \( F^1_{p,\varepsilon|F_0} \) 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 (19) 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.