Submitted to *Operations Research*
manuscript (Please, provide the manuscript number!)

A Composite Risk Measure Framework for Decision Making under Uncertainty

Pengyu Qian
School of Mathematical Sciences, Peking University, China, pengyu.qian@pku.edu.cn

Zizhuo Wang
Department of Industrial and Systems Engineering, University of Minnesota, Minneapolis, MN 55455, zwang@umn.edu

Zaiwen Wen
Beijing International Center for Mathematical Research, Peking University, China, wenzw@math.pku.edu.cn

In this paper, we present a unified framework for decision making under uncertainty. Our framework is based on the composite of two risk measures, where the inner risk measure accounts for the risk of decision given the exact distribution of uncertain model parameters, and the outer risk measure quantifies the risk that occurs when estimating the parameters of distribution. We show that the model is tractable under mild conditions. The framework is a generalization of several existing models, including stochastic programming, robust optimization, distributionally robust optimization, etc. Using this framework, we study a few new models which imply probabilistic guarantees for solutions and yield less conservative results comparing to traditional models. Numerical experiments are performed on portfolio selection problems to demonstrate the strength of our models.

1. Introduction

In this paper, we consider a decision maker who wants to minimize an objective function $H(\mathbf{x}, \boldsymbol{\xi})$, where $\mathbf{x} \in \mathbb{R}^n$ is the decision variable and $\boldsymbol{\xi} \in \mathbb{R}^s$ is some uncertain/unknown parameter related to the model. For example, in a newsvendor problem, $\mathbf{x}$ is the order amount of newspapers by a newsvendor, and $\boldsymbol{\xi}$ is the uncertain future demand. Similarly, in an investment problem, $\mathbf{x}$ is the portfolio chosen by a portfolio manager, and $\boldsymbol{\xi}$ is the unknown future returns of the instruments.

The existence of the uncertain parameters distinguishes the problem from ordinary optimization problems and has led to several decision making paradigms.

One of the earliest attempts to deal with such decision making problems under uncertainty was proposed by Dantzig (1955), where it was assumed that the distribution of $\boldsymbol{\xi}$ is known exactly and the decision is chosen to minimize the expectation of $H(\mathbf{x}, \boldsymbol{\xi})$. Such an approach is called stochastic programming. Another approach named robust optimization initiated by Soyster (1973) supposes that all possible values of $\boldsymbol{\xi}$ lie within an uncertainty set, and the decision should be made to minimize the worst-case value of $H(\mathbf{x}, \boldsymbol{\xi})$. Stochastic programming and robust optimization models can be viewed as two extremes in the spectrum of available information in decision making under uncertainty. There are models in between these two extremes. For example, distributionally robust models (Scarf et al. 1958, Dupačová 1987, Delage and Ye 2010, etc.) take into account both the stochastic and the robust aspects, where the distribution of $\boldsymbol{\xi}$ is assumed to belong to a certain distribution set and the worst-case expectation of $H(\mathbf{x}, \boldsymbol{\xi})$ is minimized. There are also various models which minimize certain risk (Rockafellar and Uryasev 2000, Gaivoronski and Pflug 2005, etc.) or the worst-case risk (El Ghaoui et al. 2003, Zhu and Fukushima 2009, etc.) of $H(\mathbf{x}, \boldsymbol{\xi})$. We will present a more detailed review of these models in Section 2. In addition to the study of individual models, there have been recent efforts to seek connections between different models and to put forward more general models. For instance, Bertsimas and Brown (2009) and Natarajan et al. (2009) show that uncertainty sets can be constructed according to decision maker's risk preference, Bertsimas et al. (2014) propose a general framework for data-driven robust optimization, and Wiesemann et al. (2014) propose a framework for distributionally robust models.

While some models have demonstrated their effectiveness in practice, there are still some ignored issues in the existing literature:

1. There lacks a unified framework which includes all the models above, namely, stochastic programming model, robust optimization model, distributionally robust model and worst-case risk models.

2. Though risk measures have been imposed on the objective function to deal with parameter uncertainty (Artzner et al. 1999, Bertsimas and Brown 2009, etc.), no attempt has been made to impose risk measure on the expectation or other functionals of the objective function with respect to distributional uncertainty.

3. Bayesian approach has not been fully considered when modeling decision making under distributional uncertainty in the existing literature, albeit its appropriateness for such problems.

The goal of our paper is to fill these gaps and build a unified modeling framework for decision making under uncertainty. Our unified framework is based on a risk measure interpretation for robustness and encompasses several popular models, such as stochastic programming, robust optimization, distributionally robust optimization, worst-case risk models, etc. Specifically, we minimize the composite of two risk measures, where the inner risk measure accounts for the risk of decision given the exact distribution of parameters, and the outer risk measure quantifies the risk that occurs when estimating the parameters of distribution. For the outer risk measure, we take a Bayesian approach and consider the posterior distribution of distribution parameters. We show that the composite of risk measures is convex as long as both the inner and the outer risk measures are convex risk measures. We also use this framework to construct several new models which have real world meanings and perform numerical tests on these models which demonstrate their strength.

We summarize our contributions as follows:

1. We propose a composite risk measure (CRM) framework for decision making under uncertainty where the composite of two risk measures is minimized. It is a generalization of several existing models. We show that the corresponding optimization problem is convex under mild conditions.

2. We take a novel approach to deal with distributional uncertainty by making use of risk measure and Bayesian posterior distribution of distribution parameters.

3. Using the composite risk measure framework, we study a VaR-Expectation model, a CVaR-Expectation model and a CVaR-CVaR model, investigating their tractability and probabilistic

guarantees. Numerical experiments show that these models can be solved efficiently and perform well in portfolio selection problems.

The remainder of this paper is organized as follows. In Section 2, we briefly review the existing models for decision making under uncertainty. In Section 3, we present our composite risk measure framework and show that several existing models fall into the framework. Several new models within the general framework are proposed in Section 4. Numerical results for these new models are shown in Section 5.

Notations. Throughout the paper, the following notations will be used. Ordinary lower case letters $(x, y, \dots)$ denote scalars, boldfaced lower case letters $(\mathbf{x}, \mathbf{y}, \dots)$ denote vectors. Specifically, $\mathbf{x} \in \mathbb{R}^n$ denotes the decision variable and $\boldsymbol{\xi} \in \mathbb{R}^s$ denotes the uncertain/unknown parameters. $H(\mathbf{x}, \boldsymbol{\xi})$ is the loss function under decision $\mathbf{x}$ and parameter $\boldsymbol{\xi}$.

2. Review of Existing Models for Decision Making under Uncertainty

There have been many models proposed in the literature that study decision making problem under uncertainty. In this section, we provide a review of those existing models.

2.1. Stochastic Programming

One of the most popular ways to solve decision making problem under uncertainty is through stochastic programming. In stochastic programming models, one assumes that the full distributional information of $\boldsymbol{\xi}$ is available. Then to choose an optimal decision variable $\mathbf{x}$, one considers a certain functional of the random loss $H(\mathbf{x}, \boldsymbol{\xi})$. Examples of this approach include:

Expectation optimization. Dantzig (1955) considers the case where the objective is to minimize the expectation of the random loss $H(\mathbf{x}, \boldsymbol{\xi})$. Namely, the optimization problem is

$$\min_{\mathbf{x}} \ \mathbb{E}[H(\mathbf{x}, \boldsymbol{\xi})]. \tag{1}$$

There are much literature that study such optimization problems. We refer interested readers to Shapiro et al. (2009) for a comprehensive review.

Value-at-Risk (VaR) optimization. VaR has enjoyed great popularity as a measure of downside risk in finance industry since its introduction in the 1990s (see RiskMetrics 1996). For fixed $\mathbf{x}$, the δ-VaR of $H(\mathbf{x}, \boldsymbol{\xi})$ is defined as the δ quantile of the random loss $H(\mathbf{x}, \boldsymbol{\xi})$. Mathematically,

$$\mathrm{VaR}_\delta(H(\mathbf{x}, \boldsymbol{\xi})) \triangleq \inf\{t \in \mathbb{R} | \mathbb{P}(H(\mathbf{x}, \boldsymbol{\xi}) \geqslant t) \leqslant 1 - \delta\}. \tag{2}$$

The corresponding VaR optimization problem can be written as:

$$\begin{aligned} \min_{\mathbf{x}, t} \quad & t \\ \text{s.t.} \quad & \mathbb{P}(H(\mathbf{x}, \boldsymbol{\xi}) \geqslant t) \leqslant 1 - \delta. \end{aligned} \tag{3}$$

The VaR optimization problem has been studied extensively in the literature, see, e.g., Kast et al. (1998), Lucas and Klaassen (1998), etc. However, there are three main drawbacks of VaR. First, it doesn't take into account the magnitude of loss beyond VaR, resulting in decision maker's preference for taking 'excessive but remote' risks (Einhorn and Brown 2008). Second, it is not subadditive (Artzner et al. 1999), meaning that the VaR of a combined portfolio can be larger than the sum of the VaRs of its components, a property that is not desired in many applications. Lastly, $\mathrm{VaR}_\delta(H(\mathbf{x}, \boldsymbol{\xi}))$ is usually not convex in $\mathbf{x}$ (Birge and Louveaux 2011), making the optimization problem intractable.

Conditional Value-at-Risk (CVaR) optimization. To overcome the drawbacks of VaR, researchers further proposed a modified version of VaR, the CVaR (also called the expected shortfall in some literature). For a certain random loss X, the δ-CVaR is the expected loss in the worst $1 - \delta$ cases. Mathematically, it can be written as:

$$\mathrm{CVaR}_\delta(X) = \frac{1}{1-\delta} \int_\delta^1 \mathrm{VaR}_s(X) ds.$$

For atomless distributions, CVaR is equivalent to the conditional expectation of the loss beyond VaR, namely,

$$\mathrm{CVaR}_\delta(X) = \mathbb{E}[X | X \geqslant \mathrm{VaR}_\delta(X)].$$

Rockafellar and Uryasev (2000) show that CVaR can be obtained by solving the following convex program:

$$\min_{\alpha\in\mathbb{R}} \ \alpha + \frac{1}{1-\delta}\mathbb{E}[(H(\mathbf{x},\boldsymbol{\xi})-\alpha)^+], \tag{4}$$

which leads to another definition of CVaR. This formulation brings computational convenience as the objective function is explicit and convex in α.

2.2. Robust Optimization

Another popular way for decision making under uncertainty is to use a robust optimization approach. In the robust optimization approach, instead of assuming the distributional knowledge of $\boldsymbol{\xi}$, one only assumes that $\boldsymbol{\xi}$ takes values in a certain uncertainty set Ξ. Then, when choosing the decision variable, one considers the worst-case outcome associated with each decision, where the worst-case scenario is chosen from the specified uncertainty set. The optimization problem can be written as follows:

$$\min_{\mathbf{x}} \ \max_{\boldsymbol{\xi}\in\Xi} \ H(\mathbf{x},\boldsymbol{\xi}). \tag{5}$$

Robust optimization problems have more general forms where constraints instead of objective function are affected by parameter uncertainty (e.g. Bertsimas et al. 2011). Robust optimization was first proposed by Soyster (1973) and has attracted much attention in the past few decades. For a comprehensive review of the literature, we refer readers to Bertsimas et al. (2011) and Ben-Tal et al. (2009).

Choosing a suitable uncertainty set is essential in formulating a robust optimization problem. Two main issues should be taken into consideration when designing uncertainty sets:

Tractability. Only a few uncertainty sets will lead to tractable counterparts for the original problem. Some known cases include polyhedral uncertainty set (Ben-Tal and Nemirovski 1999), ellipsoidal uncertainty set (Ben-Tal and Nemirovski 1999, El Ghaoui and Lebret 1997 and El Ghaoui et al. 1998), norm uncertainty set (Bertsimas et al. 2004), etc.

Conservativeness. Intuitively, when the uncertainty set is very large, the resulting decision will be very robust, but sometimes too conservative to be of practical use. Much work has been done on the choice of uncertainty set, see Bertsimas and Sim (2004), Chen et al. (2007) and Bertsimas et al. (2014).

2.3. Distributionally Robust Models

In practice, it is often the case that one only has partial information about the distribution of $\boldsymbol{\xi}$, such as first and second moments. Applying stochastic programming in those cases is not feasible. Meanwhile, using a robust optimization approach is not straightforward either and often results in overly-conservative solutions. As a result, an intermediate path has been proposed which is called the distributionally robust optimization (DRO) model. In the DRO models, the decision maker constructs an uncertainty set $\mathcal{F}$ of the underling distribution and minimizes the expected value of the objective function under the worst-case distribution chosen from $\mathcal{F}$. That is, one considers the following problem:

$$\min_{\mathbf{x}} \max_{F\in\mathcal{F}} \mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x},\boldsymbol{\xi})], \tag{6}$$

where $\mathbb{E}_{\boldsymbol{\xi}\sim F}[\cdot]$ denotes the expectation of $H(\mathbf{x},\boldsymbol{\xi})$ when F is the distribution of $\boldsymbol{\xi}$.

Distributionally robust model was first proposed in Scarf et al. (1958). After its introduction, several different choices of the uncertainty set $\mathcal{F}$ have been proposed. We will review two most discussed types of uncertainty sets below and refer readers to Wiesemann et al. (2014) and references therein for other types of uncertainty sets.

Distributions with partial knowledge on moments. Scarf et al. (1958), Dupačová (1987), Prékopa (1995), Bertsimas and Popescu (2005), etc. consider a family of distributions with known moments. Specifically, Scarf et al. (1958) consider the following optimization problem (in the context of an inventory problem):

$$\min_{x\in\mathbb{R}} \max_{F\in\mathcal{F}(\mu_0,\sigma_0)} \mathbb{E}_{\xi\sim F}[c(x-\xi)^+ + r(\xi-x)^+]$$

$$\text{where } \mathcal{F}(\mu_0,\sigma_0)=\left\{ F(\xi)\in\mathcal{M} \left| \begin{array}{l} \mathbb{P}(\xi\in\mathbb{R}^+)=1 \\ \mathbb{E}_{\xi\sim F}(\xi)=\mu_0 \\ \mathbb{E}_{\xi\sim F}(\xi-\mu_0)^2=\sigma_0^2 \end{array} \right. \right\}. \tag{7}$$

In (7), $\mathcal{M}$ is the set of all probability measures in the probability space where ξ is defined and c and r are given parameters. Scarf et al. (1958) show that the worst-case distribution is a two-point distribution with given decision x and derive a closed-form solution to (7).

Delage and Ye (2010) consider a more general form of moment constraints. Namely, the uncertainty set is constructed as:

$$\mathcal{F}(\Xi,\boldsymbol{\mu}_0,\Sigma_0,\gamma_1,\gamma_2)=\left\{ F(\boldsymbol{\xi})\in\mathcal{M} \left| \begin{array}{l} \mathbb{P}(\boldsymbol{\xi}\in\Xi)=1 \\ (\mathbb{E}_{\boldsymbol{\xi}\sim F}(\boldsymbol{\xi})-\boldsymbol{\mu}_0)^T\Sigma_0^{-1}(\mathbb{E}_{\boldsymbol{\xi}\sim F}(\boldsymbol{\xi})-\boldsymbol{\mu}_0)\leqslant\gamma_1 \\ \mathbb{E}_{\boldsymbol{\xi}\sim F}[(\boldsymbol{\xi}-\boldsymbol{\mu}_0)(\boldsymbol{\xi}-\boldsymbol{\mu}_0)^T]\preceq\gamma_2\Sigma_0 \end{array} \right. \right\}. \tag{8}$$

Delage and Ye (2010) show that problem (8) can be solved by a convex optimization problem. They also provide a data-driven method for choosing the parameters γ_1 and γ_2.

More recently, Wiesemann et al. (2014) consider a general model which incorporates both moment information and support information. The distributional uncertainty set considered is:

$$\mathcal{F}=\left\{ F\in\mathcal{M}(\mathbb{R}^l\times\mathbb{R}^m) \left| \begin{array}{l} \mathbb{E}_{\boldsymbol{\xi}\sim F}(A\boldsymbol{\xi}+B\boldsymbol{\eta})=b \\ \mathbb{P}_{\boldsymbol{\xi}\sim F}[(\boldsymbol{\xi},\boldsymbol{\eta})\in\mathcal{C}_i]\in[\bar{\mathbf{p}}_i,\underline{\mathbf{p}}_i],\forall i\in\mathcal{I} \end{array} \right. \right\}, \tag{9}$$

where $\mathcal{M}(\mathbb{R}^l\times\mathbb{R}^m)$ represents probability distribution on $\mathbb{R}^l\times\mathbb{R}^m$, $\boldsymbol{\xi}\in\mathbb{R}^l$ is the random term, $\boldsymbol{\eta}\in\mathbb{R}^m$ is an auxiliary random vector, $A\in\mathbb{R}^{k\times l}$, $B\in\mathbb{R}^{k\times m}$, $b\in\mathbb{R}^k$ are predetermined parameters and $\mathcal{C}_i$ are conic confidence sets. It is shown that when certain conditions are satisfied, model (6) with distributional uncertainty set (9) can be solved by conic programming.

Distributions related to a given distribution. Ben-Tal et al. (2013) consider a set of distributions that arise from ϕ-divergence. Namely,

$$\mathcal{F}=\left\{ f\in\mathbb{R}^m \left| I_\phi(f,\hat{f})\leqslant\rho,\sum_{i=1}^m f_i=1,f_i\geqslant 0,i=1,2,\ldots,m \right. \right\}, \tag{10}$$

where $\hat{f}$ is a given probability vector (e.g., the empirical distribution vector). $I_\phi(f, \hat{f})$ is the ϕ-divergence between two probability vectors f and $\hat{f}$ (Pardo 2005). Ben-Tal et al. (2013) show that the robust linear constraint

$$(\mathbf{a} + \mathbb{E}_{\boldsymbol{\xi}\sim F}(\boldsymbol{\xi}))^T \mathbf{x} \leqslant \beta, \quad \forall F \in \mathcal{F},$$

can be written equivalently as a linear constraint, a conic constraint, or a convex constraint, depending on the choice of ϕ, where $\mathbf{a} \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ are fixed parameters, $\mathbf{x} \in \mathbb{R}^n$ is the decision variable and $\mathcal{F}$ is defined in (10). Similar approaches have been taken by Wang et al. (2013) and Klabjan et al. (2013).

Bertsimas et al. (2014) propose a model where the distributional uncertainty set is constructed by means of hypothesis test given a set of available data. Namely, two hypotheses are compared:

$$H_0 : \mathbf{p}^* = \mathbf{p}_0 \quad vs. \quad H_A : \mathbf{p}^* \neq \mathbf{p}_0,$$

where H_0 is the null hypothesis, H_A is the alternative hypothesis and $\mathbf{p}_0$ is an arbitrary distribution. By specifying a hypothesis test, e.g. χ^2-test, G-test, etc., and a confidence level ϵ, one can construct a distribution set $\mathcal{F}$ containing all the distributions that pass the test under the given set of data.

2.4. Other Choices of Objectives

In addition to the models mentioned above, there are several other models for decision making under uncertainty that have been studied in the literature.

Minimizing worst-case VaR. El Ghaoui et al. (2003) consider the problem of minimizing VaR (defined in (2)) over a portfolio of random loss $\boldsymbol{\xi}$, where only partial knowledge about the distribution F of $\boldsymbol{\xi}$ is known. Mathematically, the optimization problem is:

$$\begin{aligned} \min_{\mathbf{x},t} \quad & t \\ \text{s.t.} \quad & \mathbb{P}_{\boldsymbol{\xi}\sim F}(\boldsymbol{\xi}^T \mathbf{x} \geqslant t) \leqslant 1 - \delta, \quad \forall F \in \mathcal{F} \\ & \mathbf{x} \in \mathcal{X}, \end{aligned} \tag{11}$$

where $\mathcal{F}$ is the set of all probability measures with given first two moments μ_0 and Σ_0 ($\Sigma_0 \succeq 0$). Using the exact Chebyshev bound given in Bertsimas and Popescu (2005), El Ghaoui et al. (2003) show that problem (11) can be reduced to a second-order cone program (SOCP). El Ghaoui et al. (2003) further show that if $\mathcal{F}$ is a set of distributions with the first and second moments (μ, Σ) satisfying $(\mu, \Sigma) \in \text{conv}\{(\mu_1, \Sigma_1), (\mu_2, \Sigma_2), \ldots, (\mu_k, \Sigma_k)\}$ or (μ, Σ) is bounded in a componentwise fashion, problem (11) can be solved by an SOCP or a semi-definite program (SDP) respectively.

Minimizing worst-case CVaR Zhu and Fukushima (2009) solve the problem of optimizing the CVaR (defined in (4)) of a portfolio when the distribution F of the random return $\boldsymbol{\xi}$ is only known to belong to an uncertainty set $\mathcal{F}$ instead of being exactly known, namely,

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{F\in\mathcal{F}} \text{CVaR}_{F,\delta}[H(\mathbf{x}, \boldsymbol{\xi})], \tag{12}$$

where $\text{CVaR}_{F,\delta}(\cdot)$ denote the δ-CVaR of a random variable whose distribution is F. Zhu and Fukushima (2009) show that for certain forms of $\mathcal{F}$, problem (12) can be transformed to a convex optimization problem.

3. A Composite Risk Measure Framework

In this section, we present a unified framework for decision making problem under uncertainty. Our framework encompasses all the decision paradigms discussed in Section 2 and can be used to generate new ones.

The idea of our framework is based on risk measures defined as follows:

Definition 1. Let $\mathcal{L}$ be a set of random variables defined on the sample space Ω. A functional $\rho(\cdot) : \mathcal{L} \to \mathbb{R}$ is a risk measure if it satisfies the following properties:

1. *Monotonicity:* For any $\mathbf{X}, \mathbf{Y} \in \mathcal{L}$, if $\mathbf{X} \geqslant \mathbf{Y}$, then $\rho(\mathbf{X}) \geqslant \rho(\mathbf{Y})$, where $\mathbf{X} \geqslant \mathbf{Y}$ means that $\mathbf{X}(\omega) \geqslant \mathbf{Y}(\omega)$ for any $\omega \in \Omega$.

2. *Translation invariance:* For any $\mathbf{X} \in \mathcal{L}$ and $c \in \mathbb{R}$, $\rho(\mathbf{X} + c) = \rho(\mathbf{X}) + c$.

In addition, Artzner et al. (1999) define a subset of risk measures satisfying some structural properties presented as follows.

DEFINITION 2. If a risk measure $\rho(\cdot)$ satisfies the following properties:

1. *Convexity:* For any $\mathbf{X}, \mathbf{Y} \in \mathcal{L}$ and $\lambda \in [0,1]$, $\rho(\lambda\mathbf{X} + (1-\lambda)\mathbf{Y}) \leqslant \lambda\rho(\mathbf{X}) + (1-\lambda)\rho(\mathbf{Y})$;
2. *Positive homogeneity:* For any $\mathbf{X} \in \mathcal{L}$ and $\lambda \geqslant 0$, $\rho(\lambda\mathbf{X}) = \lambda\rho(\mathbf{X})$,

then $\rho(\cdot)$ is called a coherent risk measure. If only convexity holds, it is called a convex risk measure (Föllmer and Schied 2002).

To establish our framework, first we note that given $\mathbf{x}$, $H(\mathbf{x}, \boldsymbol{\xi})$ is a random variable defined on the sample space of $\boldsymbol{\xi}$ (which we denote by Ω_0). We denote such a random variable by $Y(\mathbf{x})$. Define $g_F(\cdot)$ to be a risk measure for $Y(\mathbf{x})$, where the subscript F is used to show the dependence of this risk measure on the choice of distribution F. Now we further define a measurable space for F: (Ω_1, Σ_1), where Ω_1 denotes the space of all the distribution functions for $\boldsymbol{\xi}$, and Σ_1 is a σ-algebra defined on the space of such distributions. Moreover, we can define a measure $\mathbb{P}_1$ on such a space using concepts from Bayesian statistics (see the following passage for detailed discussions). With this definition, the risk measure $g_F(Y(\mathbf{x}))$ can be viewed as a random variable too in the following way:

$$Z(\mathbf{x}) : F \in \Omega_1 \rightarrow g_F(Y(\mathbf{x})) \in \mathbb{R}. \tag{13}$$

We denote the linear space of $Z(\mathbf{x})$ by $\mathcal{Z}$. Finally, since $Z(\mathbf{x})$ is a random variable, we can apply another risk measure $\mu : \mathcal{Z} \rightarrow \mathbb{R}$ and consider the following optimization problem:

$$\min_{\mathbf{x} \in \mathcal{X}} \ \mu(Z(\mathbf{x})). \tag{14}$$

Therefore, our framework can be written as follows:

$$\min_{\mathbf{x} \in \mathcal{X}} \ \mu(g_F(H(\mathbf{x}, \boldsymbol{\xi}))). \tag{15}$$

We call this the composite-risk-measure (CRM) framework for decision making under uncertainty. In the following discussions, we will refer to $g_F(\cdot)$ as the inner risk measure and $\mu(\cdot)$ as the outer risk measure. We first present the following tractability result for problem (15).

PROPOSITION 1. *Optimization problem (15) is a convex optimization problem if the following holds:*

1. $H(\mathbf{x}, \boldsymbol{\xi})$ *is convex in* $\mathbf{x}$*;*
2. $\mathcal{X}$ *is convex;*
3. $\mu(\cdot)$ *is a convex risk measure;*
4. $g_F(\cdot)$ *is a convex risk measure for each* $F \in \mathcal{F}$*.*

Proof. We show that $\mu(g_F(H(\mathbf{x}, \boldsymbol{\xi})))$ is a convex function of $\mathbf{x}$. For any $\mathbf{x}, \mathbf{y} \in \mathcal{X}$ and $0 \leqslant \lambda \leqslant 1$, we have:

$$\begin{aligned}\lambda\mu(g_F(H(\mathbf{x}, \boldsymbol{\xi}))) + (1-\lambda)\mu(g_F(H(\mathbf{y}, \boldsymbol{\xi}))) &\geqslant \mu(\lambda g_F(H(\mathbf{x}, \boldsymbol{\xi})) + (1-\lambda) g_F(H(\mathbf{y}, \boldsymbol{\xi}))) \\ &\geqslant \mu(g_F(\lambda H(\mathbf{x}, \boldsymbol{\xi}) + (1-\lambda) H(\mathbf{y}, \boldsymbol{\xi}))) \\ &\geqslant \mu(g_F(H(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}, \boldsymbol{\xi}))),\end{aligned}$$

where the first line follows from the convexity of $\mu(\cdot)$; the second line follows from the convexity of $g_F(\cdot)$ and the monotonicity of $\mu(\cdot)$; the third line follows from the convexity of $H(\cdot, \boldsymbol{\xi})$ and the monotonicity of $g_F(\cdot)$ and $\mu(\cdot)$. □

Now we turn to the distribution $\mathbb{P}_1$ over Ω_1. We make the following assumption in our discussion:

ASSUMPTION 1. Ω_1 *is parameterized by a finite number of parameters.*

In fact, Assumption 1 does not cause much loss of generality. Many distribution families we are interested in are parameterized by a finite number of parameters. For example, if $\mathcal{F}$ is the family of discrete distributions whose probability mass function is: $\mathbb{P}(\boldsymbol{\xi} = \boldsymbol{\xi}_i) = p_i, i = 1, ..., m$, then

$$\Omega_1 = \left\{ (p_1, p_2, \ldots, p_m) \in \mathbb{R}^m \,\middle|\, \sum_{i=1}^{m} p_i = 1; p_i \geqslant 0, \forall i = 1, 2, \ldots, m \right\}$$

For family of multivariate normal distributions:

$$dF(\boldsymbol{\xi}) = \frac{1}{\sqrt{(2\pi)^s |\Sigma|}} \exp(-\frac{1}{2}(\boldsymbol{\xi} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{\xi} - \boldsymbol{\mu})),$$

we have

$$\Omega_1 = \left\{ (\boldsymbol{\mu}, \Sigma) | \boldsymbol{\mu} \in \mathbb{R}^s, \Sigma \in \mathcal{S}_+^s \right\},$$

where $\mathcal{S}_+^s$ is the cone of positive semi-definite matrices. For family of mixture distributions $dF(\boldsymbol{\xi}) = \sum_{i=1}^m \lambda_i p^i(\cdot) : \sum_{i=1}^m \lambda_i = 1, \lambda_i \geqslant 0, i = 1, 2, ..., m$, where $p^i(\cdot), i = 1, 2, ..., m$, are predetermined distributions, we have

$$\Omega_1 = \left\{ (\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathbb{R}^m \left| \sum_{i=1}^m \lambda_i = 1; \lambda_i \geqslant 0, \forall i = 1, 2, \ldots, m \right. \right\}.$$

In the framework of Bayesian statistics, observations of $\boldsymbol{\xi}$ are treated as given rather than samples randomly drawn from an underlying distribution. Meanwhile, the parameters of distribution are handled as random variables which reflect the likelihood it takes each value given the observations. We denote the distribution parameters as $\Theta = (\theta_1, \ldots, \theta_m) \in \Omega_1$. To obtain the distribution of Θ, one should first specify a prior distribution $f(\Theta)$ which expresses one's belief about Θ when no observation is available. Prior distribution can be either informative or uninformative. Then, when observed data $\Xi = (\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_N)$ are collected, one can derive the posterior distribution $p(\Theta|\Xi)$ using Bayes' formula:

$$p(\Theta|\Xi) = \frac{f(\Theta) \prod_{i=1}^N L(\boldsymbol{\xi}_i|\Theta)}{\int_{\Omega_1} f(\Theta) \prod_{i=1}^N L(\boldsymbol{\xi}_i|\Theta) d\Theta}, \tag{16}$$

where $p(\Theta|\Xi)$ is the posterior distribution given data Ξ, $\prod_{i=1}^N L(\boldsymbol{\xi}_i|\Theta)$ is the likelihood function and $\int_{\Omega_1} f(\Theta) \prod_{i=1}^N L(\boldsymbol{\xi}_i|\Theta) d\Theta$ is the normalizing factor. Note that the integration in (16) should be replaced by summation when a discrete distribution instead of a continuous distribution is considered. In practice, if one wants to sample from the posterior distribution in (16), according to the Metropolis-Hastings algorithm, one only needs to be able to compute the numerator, which is usually easy to do. In this way, the distribution $\mathbb{P}_1$ over Ω_1 can be defined and sampled from easily, and it is often the case that such definitions are data-driven.

3.1. Relation to the Models in Section 2

In the following, we show that all the optimization models discussed in Section 2 can be viewed as special cases of our proposed composite risk measure framework.

Stochastic programming. In stochastic programming models, we assumed that we know the distribution F_0 exactly, namely, $\mathcal{F} = \{F_0\}$. Therefore, models (1), (3) and (4) can be viewed as the outer risk measure taken to be a singleton:

$$\mu(g_F(\cdot)) = g_{F_0}(\cdot),$$

and the inner risk measures are chosen to be expectation, VaR and CVaR, respectively.

Distributionally robust optimization. In the distributionally robust optimization models, the inner measure is chosen to be the expectation measure, while the outer measure can be viewed as the worst-case risk measure:

$$\mathrm{WC}(\mathbf{Z}) = \inf\{\alpha \,|\mathbb{P}(\mathbf{Z} \leqslant \alpha) = 1\}, \tag{17}$$

where $\mathbf{Z}$ is a random variable defined on $(\Omega_1, \Sigma_1, \mathbb{P}_1)$ with Ω_1 chosen to be the distribution set $\mathcal{F}$.

In fact, distributionally robust model also covers the singleton-coherent risk measure model:

$$\min_{\mathbf{x}} \ \mu(\mathbf{Y}(\mathbf{x})).$$

Artzner et al. (1999) show that coherent risk measures are closely related to worst-case risk measures. The exact relationship is given in the following theorem (we use $\ll$ to denote absolute continuity between probability measures).

THEOREM 1. *Let $\mathcal{X}$ be any linear space of random variables defined on a probability space $(\Omega, \Sigma, \mathbb{P})$. A functional $\rho(\cdot) : \mathcal{X} \to \mathbb{R}$ is a coherent risk measure if and only if there exists a family of probability measures $\mathcal{Q}$ with $\mathbb{Q} \ll \mathbb{P}$ for all $\mathbb{Q} \in \mathcal{Q}$ such that*

$$\rho(\mathbf{X}) = \sup_{\mathbb{Q} \in \mathcal{Q}} \ \mathbb{E}_{\mathbb{Q}}(\mathbf{X}), \quad \forall \mathbf{X} \in \mathcal{X},$$

where $\mathbb{E}_{\mathbb{Q}}(\mathbf{X})$ denotes the expectation of the random variable $\mathbf{X}$ under the measure $\mathbb{Q}$ (as opposed to the measure of $\mathbf{X}$ itself).

For example, for $\text{CVaR}_\delta(\cdot)$, the corresponding set of distribution is $\mathcal{Q} = \{\mathbb{Q} \ll \mathbb{P} \,|d\mathbb{Q}/d\mathbb{P} \leqslant (1-\delta)^{-1}\}$. The above theorem shows that coherent risk measures can be represented by the worst-case expectations taken over a set of probability distributions. The proof of Theorem 1 in fact predates the introduction of coherent risk measure, see Huber (1981). And this result has been used in several recent works that study distributionally robust optimizations, see, e.g., Bertsimas and Brown (2009) and Natarajan et al. (2009).

Robust optimization. By using the distribution set $\mathcal{F}$ where each $F \in \mathcal{F}$ is a distribution putting all its weight on one point $\boldsymbol{\xi} \in \boldsymbol{\Xi}$, the distributionally robust model (6) reduces to a robust optimization model (5) which falls into our framework.

Worst-Case CVaR and VaR optimization. Comparing model (11) and (12) to the unified model (15), we see that the corresponding inner risk measures are VaR and CVaR, respectively, while the outer risk measure is the worst-case risk measure.

By choosing different combinations of outer and inner risk measures, one can come up with more optimization models. However, some of those models reduce to the models above after transformation. Some examples of these models are presented as follows.

Two-fold expectation as expectation. Choosing outer risk measure $\mu(\cdot)$ and inner risk measure $g_F(\cdot)$ both as expectations, we obtain the following optimization model:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathbb{E}(\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x},\boldsymbol{\xi})]). \tag{18}$$

Denote the cumulative distribution function of the random term $\boldsymbol{\xi}$ by $F_\Theta(\boldsymbol{\xi})$, we have

$$\begin{aligned}
\mathbb{E}(\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x},\boldsymbol{\xi})]) &= \int_{\Omega_1}\int_{\mathbb{R}^s} H(\mathbf{x},\boldsymbol{\xi}) dF_\Theta(\boldsymbol{\xi})\ p(\Theta)d\Theta \\
&= \int_{\mathbb{R}^s} H(\mathbf{x},\boldsymbol{\xi}) \left[\int_{\Omega_1} p(\Theta) dF_\Theta(\boldsymbol{\xi}) d\Theta\right] \\
&= \int_{\mathbb{R}^s} H(\mathbf{x},\boldsymbol{\xi}) d\hat{F}_\Theta(\boldsymbol{\xi}) \\
&= \hat{\mathbb{E}}[H(\mathbf{x},\boldsymbol{\xi})],
\end{aligned}$$

Table 1 Different composites of risk measures

$\mu(\cdot)$ \ $g_F(\cdot)$	Expectation	CVaR	VaR	Worst-Case
Singleton	(1)	(4)	(3)	(5)
Expectation	$\sim$(1)	$\times$	$\times$	$\times$
VaR	$\times$	$\times$	$\times$	$\times$
Worst-Case	(6)	(12)	(11)	$\sim$(5)

where $d\hat{F}_\Theta = \int_{\Omega_1} p(\Theta) dF_\Theta d\Theta$ can be viewed as the expected measure (e.g., if Ω_1 contains only continuous distributions, then this is just a weighted average over all the density functions). Therefore, problem (18) is equivalent to the expectation optimization problem:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \hat{\mathbb{E}}[H(\mathbf{x},\boldsymbol{\xi})].$$

Two-fold worst-case as worst-case. Choosing outer risk measure $\mu(\cdot)$ and inner risk measure $g_F(\cdot)$ both as worst-case risk measures, we obtain the following optimization model:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \max_{F\in\mathcal{F}} \ \max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}_F} H(\mathbf{x},\boldsymbol{\xi}), \tag{19}$$

where $\boldsymbol{\Xi}_F$ is the uncertainty set for $\boldsymbol{\xi}$ when the distribution of $\boldsymbol{\xi}$ is F. This problem can be reduced to model (5):

$$\min_{\mathbf{x}\in\mathcal{X}} \ \max_{\boldsymbol{\xi}\in\boldsymbol{\Xi}} H(\mathbf{x},\boldsymbol{\xi}), \tag{20}$$

where $\Xi = \{\boldsymbol{\xi} : \exists F_0 \in \mathcal{F}, \boldsymbol{\xi} \in \Xi_{F_0}\}$. Example of this model can be found in Bertsimas et al. (2014). All the cases discussed in this section are summarized in Table 1, where the symbol $\sim$ means that the model is equivalent to another model labeled by the number, the symbol $\times$ means that such combination of risk measures has not been considered yet.

4. Constructing New Models

In this section, we use our framework to propose and study a few new paradigms for decision making under uncertainty. In the following, we continue to use the notation Ω_1 to denote the sample space of F and $\mathbb{P}_1$ to denote the probability distribution over Ω_1. As we have discussed earlier, $\mathbb{P}_1$ can be derived as a posterior distribution from a Bayesian approach.

4.1. Minimizing VaR-Expectation

When minimizing the expectation of random loss $H(\mathbf{x}, \boldsymbol{\xi})$ under distributional uncertainty, we can choose the outer risk measure as δ-VaR to provide a probabilistic guarantee:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathrm{VaR}_\delta(\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})]), \tag{21}$$

where $\mathcal{X}$ is the feasible set of $\mathbf{x}$. This model can be interpreted as finding a decision variable to minimize the threshold such that the chance that the expected loss exceeds the threshold is small, or in other words, this model can be viewed as minimizing the upper bound for the one-sided δ-confidence interval for the expected loss. It could be applicable in the context where the expected value is a common criterion to evaluate the loss while there is uncertainty about the underlying distribution.

Note that model (21) shares similar spirit as the distributionally robust model (6). Both models are designed to deal with parameter uncertainty. However, in distributionally robust models such as Delage and Ye (2010) and Bertsimas et al. (2014), it is assumed that there exists a true underlying distribution $\bar{F}$ of $\boldsymbol{\xi}$ and the distribution set $\mathcal{F}$ is chosen as a confidence region of $\bar{F}$ to hedge against uncertainty. And the distribution set does not depend on the decision $\mathbf{x}$. In contrast, in (21), the distribution set (for the VaR) is dependent on $\mathbf{x}$. This is often desirable since for different $\mathbf{x}$, the objective function $g_F(H(\mathbf{x}, \boldsymbol{\xi}))$ may have different properties, thus the set of unfavorable distributions may differ. As a result, solving problem (21) leads to a less conservative solution under the same robust level. To illustrate, we denote the optimal solution, the optimal value and the corresponding distribution set of problem (6) and problem (21) by $(\mathbf{x}^*_{\mathrm{DR}}, \gamma^*_{\mathrm{DR}}, \mathcal{F}_{\mathrm{DR}})$ and $(\mathbf{x}^*_{\mathrm{VaR}}, \gamma^*_{\mathrm{VaR}}, \hat{\mathcal{F}}^*_{\mathbf{x}})$ respectively, then $(\mathbf{x}^*_{\mathrm{VaR}}, \hat{\mathcal{F}}^*_{\mathbf{x}})$ is the optimal solution to the following problem:

$$\begin{aligned} \min_{\mathbf{x}\in\mathcal{X},\mathcal{F}} \ & \sup_{F\in\mathcal{F}}(\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})]) \\ \text{s.t.} \ & \mathbb{P}_1(F \notin \mathcal{F}) \leqslant 1-\delta, \end{aligned} \tag{22}$$

while $(\mathbf{x}^*_{\mathrm{DR}}, \mathcal{F}_{\mathrm{DR}})$ is only a feasible solution. Therefore, we have $\gamma^*_{\mathrm{VaR}} \leqslant \gamma^*_{\mathrm{DR}}$.

We make the following assumptions in this subsection.

Assumption 2.

1. *The loss function $H(\mathbf{x}, \boldsymbol{\xi})$ is (piecewise) continuously differentiable and convex in $\mathbf{x}$ on $\mathcal{X}$.*
2. *Both $\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})]$ and $\mathbb{E}_{\boldsymbol{\xi}\sim F}[\nabla_{\mathbf{x}} H(\mathbf{x}, \boldsymbol{\xi})]$ can be evaluated efficiently.*

The first item in Assumption 2 is necessary. Otherwise there is little hope to solve problem (21) efficiently even when the risk measures over the random loss are dropped. The second assumption ensures that standard optimization techniques can be employed to solve problem (21). When the second assumption is violated, as in the case where $\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})]$ does not have a closed-form expression and the dimension of $\mathbf{x}$ is high, one must turn to sample-based methods to evaluate $\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})]$. Consequently, the size of the problem will be large. For those cases, we will propose an approximation method in the next subsection.

We rewrite problem (21) as a chance-constrained problem:

$$\begin{aligned} \min_{t, \mathbf{x}\in\mathcal{X}} \quad & t \\ \text{s.t.} \quad & \mathbb{P}_1(\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})] \geqslant t) \leqslant 1-\delta. \end{aligned} \tag{23}$$

Thus, the optimal solution of problem (21) $\mathbf{x}^*$ represents an optimal decision where the δ quantile of the distribution of $\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x}, \boldsymbol{\xi})]$ is minimized (note that the random variable here is the distribution parameter Θ). A general approach to tackle problem (23) is the sample approximation approach (SAA) (see, e.g., Nemirovski and Shapiro 2006, Luedtke and Ahmed 2008). Using Monte Carlo method to generate N i.i.d. samples of distribution parameter $\Theta_i, i = 1, \ldots, N$ from distribution $\mathbb{P}_1$, problem (23) can be approximated by the following mixed integer nonlinear program (MINLP):

$$\begin{aligned} \min_{t, \mathbf{x}\in\mathcal{X}, z_i\in\{0,1\}} \quad & t \\ \text{s.t.} \quad & -Mz_i + \mathbb{E}_{\boldsymbol{\xi}\sim F_i}[H(\mathbf{x}, \boldsymbol{\xi})] \leqslant t, \quad i = 1, 2, ..., N \\ & \sum_{i=1}^{N} z_i = \lfloor (1-\delta)N \rfloor, \end{aligned} \tag{24}$$

where F_i is the distribution parameterized by Θ_i, and M is a large constant which should be larger than $\max_{\mathbf{x}\in\mathcal{X},i=1,\ldots,N}\mathbb{E}_{\boldsymbol{\xi}\sim F_i}[H(\mathbf{x},\boldsymbol{\xi})]-\min_{\mathbf{x}\in\mathcal{X},i=1,\ldots,N}\mathbb{E}_{\boldsymbol{\xi}\sim F_i}[H(\mathbf{x},\boldsymbol{\xi})]$. When $\mathcal{X}$ is bounded, for example, there exists a finite-valued M.

The rationale of this approximation approach is quite clear. By generating N instances of Θ, a uniform distribution over $\Theta_i, i=1,\ldots,N$ instead of $\mathbb{P}_1$ is considered when minimizing the δ-quantile of the expectation. When N is sufficiently large, problem (24) can approximate problem (23) well by the law of large numbers. Specifically, we have the following accuracy and feasibility bound which provides a guideline for choosing suitable N given desired accuracy levels. The next theorem largely follows the results in Luedtke and Ahmed (2008).

THEOREM 2. *Suppose that $\mathcal{X}$ is bounded with diameter D and $\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x},\boldsymbol{\xi})]$ is globally Lipschitz continuous with Lipschitz constant L for all $F\in\mathcal{F}$. Assume that both problems (23) and (24) are feasible and their optimal values are finite. Denote (23) and (24) by P_δ and $P_{N,\delta}$, and their optimal solutions (only the $\mathbf{x}$ part) and optimal values by $(\mathbf{x}^*_\delta,t^*_\delta)$ and $(\mathbf{x}^*_{N,\delta},t^*_{N,\delta})$, respectively. Let $0<\tau\leqslant 1-\delta$, $\epsilon\in(0,1)$, $\gamma>0$,*

$$N_0=\frac{2}{\tau^2}\left[\log\left(\frac{1}{\epsilon}\right)+n\log\left\lceil\frac{2LD}{\gamma}\right\rceil+\log\left\lceil\frac{2}{\tau}\right\rceil\right]. \tag{25}$$

Then, when $N\geqslant N_0$, with probability at least $1-2\epsilon$, we have:

$$t^*_{\delta-\tau}-\gamma\leqslant t^*_{N,\delta}\leqslant t^*_{\delta+\tau}. \tag{26}$$

Proof. When $N\geqslant N_0\geqslant\frac{1}{2\tau^2}\log\left(\frac{1}{\epsilon}\right)$, the upper bound in (26) is given by Theorem 3 in Luedtke and Ahmed (2008). Also, it follows from Theorem 10 in Luedtke and Ahmed (2008) that a feasible $(\mathbf{x},t)$ for problem:

$$\begin{aligned}\min_{t,\mathbf{x},z_i\in\{0,1\}}\quad & t \\ \text{s.t.}\quad & -Mz_i+\mathbb{E}_{\boldsymbol{\xi}\sim F_i}[H(\mathbf{x},\boldsymbol{\xi})]+\gamma\leqslant t,\quad i=1,2,...,N \\ & \sum_{i=1}^N z_i=\lfloor(1-\delta)N\rfloor\end{aligned} \tag{27}$$

is feasible for problem $P_{\delta-\tau}$ with at least $1-\epsilon$ probability. Notice that the optimal t for problem (27) equals $t^*_{N,\delta}+\gamma$. Thus, when the feasible solution $(\mathbf{x},t)$ for problem (27) is feasible for problem $P_{\delta-\tau}$, we have $t^*_{N,\delta}+\gamma \geqslant t^*_{\delta-\tau}$. This completes the proof. □

Note that in the above conditiopns, the logarithms are taken over ϵ and γ, therefore we can choose very small values of them without increasing the number of necessary samples very much. Meanwhile, experiments show that the bound in Theorem 2 is very conservative (Luedtke and Ahmed 2008) and one can choose a much smaller N in practice.

Numerically, software packages such as CPLEX (mixed integer linear problems) and MOSEK (certain MINLP problems) can be used to solve moderate sized problems. When Assumption 2 holds, problem (24) may be solved by those solvers.

REMARK 1. It is worth pointing out that model (24) can also be used to solve a family of VaR-VaR problems and VaR-CVaR problems. For portfolio optimization problems where $H(\mathbf{x},\boldsymbol{\xi})=\boldsymbol{\xi}^T\mathbf{x}$ and the distribution of $\boldsymbol{\xi}$ is a normal distribution with mean $\boldsymbol{\mu}$ and covariance Γ, both $\mathrm{VaR}_\delta(H(\mathbf{x},\boldsymbol{\xi}))$ and $\mathrm{CVaR}_\delta(H(\mathbf{x},\boldsymbol{\xi}))$ have closed-form expressions (Sarykalin et al. 2008):

$$\mathrm{VaR}_\delta(\boldsymbol{\xi}^T\mathbf{x}) = \Phi^{-1}(\delta)\sqrt{\mathbf{x}^T\Gamma\mathbf{x}}+\boldsymbol{\mu}^T\mathbf{x}, \tag{28}$$

$$\mathrm{CVaR}_\delta(\boldsymbol{\xi}^T\mathbf{x}) = \frac{1}{\sqrt{2\pi}(1-\delta)}\exp(-(\Phi^{-1}(\delta))^2/2)\sqrt{\mathbf{x}^T\Gamma\mathbf{x}}+\boldsymbol{\mu}^T\mathbf{x}, \tag{29}$$

where $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function of a standard normal distribution. In this case, we can minimize VaR_δ-VaR_ϵ and VaR_δ-CVaR_ϵ by solving the following problems:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathrm{VaR}_\delta\left(\Phi^{-1}(\epsilon)\sqrt{\mathbf{x}^T\Gamma\mathbf{x}}+\boldsymbol{\mu}^T\mathbf{x}\right), \tag{30}$$

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathrm{VaR}_\delta\left(\frac{1}{\sqrt{2\pi}(1-\epsilon)}\exp(-(\Phi^{-1}(\epsilon))^2/2)\sqrt{\mathbf{x}^T\Gamma\mathbf{x}}+\boldsymbol{\mu}^T\mathbf{x}\right), \tag{31}$$

both of which can be solved approximately by mixed integer second order cone program (MISOCP) using SAA approach.

REMARK 2. It is also worth mentioning that when the objective function $H(\mathbf{x},\boldsymbol{\xi})$ is linear in $\boldsymbol{\xi}$ (such as in the portfolio selection problem), the VaR-Expectation model is similar to a VaR model in formulation. More precisely, suppose $H(\mathbf{x},\boldsymbol{\xi})=\boldsymbol{\xi}^T f(\mathbf{x})$, then $\mathrm{VaR}_\delta(\mathbb{E}_{\boldsymbol{\xi}}[H(\mathbf{x},\boldsymbol{\xi})]) = \mathrm{VaR}_\delta((\mathbb{E}\boldsymbol{\xi})^T f(\mathbf{x}))$.

And the latter can be viewed as a VaR model with $\mathbb{E}\boldsymbol{\xi}$ as the random variable. Recently, such a formulation has been studied in Cui et al. (2013) and Wen et al. (2013) in the context of the portfolio selection problem. In particular, Wen et al. (2013) show that an alternating direction augmented Lagrangian method (ADM) can be applied to solve the problem very efficiently in that case. However, the VaR-Expectation model is different from the VaR model if the objective is not linear in $\boldsymbol{\xi}$ since the expectation can no longer be taken into $\boldsymbol{\xi}$. Moreover, even in the linear case, the VaR-Expectation model has a very different interpretation from the VaR model: In the VaR model, the uncertainty is directly in $\boldsymbol{\xi}$, and one has to assume a certain distribution of $\boldsymbol{\xi}$ (which is usually estimated from an empirical distribution); while in the VaR-Expectation model, the uncertainty is in the distribution of $\boldsymbol{\xi}$. Such different interpretations would usually lead to different uncertainty sets even with the same set of observations.

4.2. Minimizing CVaR-CVaR and CVaR-Expectation

Using the same idea as above, we formulate a robust model for CVaR optimization by choosing $\mu(\cdot)$ as δ-VaR and $g_F(\cdot)$ as ϵ-CVaR, namely,

$$\begin{aligned} &\min_{\mathbf{x}\in\mathcal{X},t\in\mathbb{R}} \quad t \\ &\text{s.t.} \quad \mathbb{P}_1(\text{CVaR}_\epsilon(H(\mathbf{x},\boldsymbol{\xi}_F)) \geqslant t) \leqslant 1-\delta, \end{aligned}$$

where $\mathbb{P}_1(\cdot)$ is the probability measure of F, and $\boldsymbol{\xi}_F$ means that the distribution of $\boldsymbol{\xi}$ is F. This model can be viewed as minimizing the upper bound for the one-sided δ-confidence interval for the CVaR.

For most cases, it is impossible to derive a closed-form expression for CVaR. In practice, sample based methods like sample average approximation (SAA) are widely used to compute CVaR. To ensure the accuracy of evaluation, the number of samples is typically large (the exact number of necessary samples will be discussed later). Therefore, if we directly replace expectation by CVaR in model (24), the resulting problem will be a large-sized mixed integer program, which is difficult

to be solved. The same situation occurs in a VaR-Expectation model where the expectation cannot be evaluated directly and must be approximated by sample average.

To deal with the issue mentioned above, we relax the outer VaR to CVaR, leading to a CVaR-CVaR (nested CVaR) model:

$$\min_{\mathbf{x}\in\mathbb{R}^n,\alpha\in\mathbb{R}} \alpha + \frac{1}{1-\delta}\mathbb{E}[(\mathrm{CVaR}_\epsilon(H(\mathbf{x},\boldsymbol{\xi}_F)) - \alpha)^+]. \tag{32}$$

Similarly, we can consider the following CVaR-Expectation model instead of the VaR-Expectation model:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathrm{CVaR}_\delta(\mathbb{E}_{\boldsymbol{\xi}\sim F}[H(\mathbf{x},\boldsymbol{\xi})]). \tag{33}$$

The reason for making the above relaxations is two fold. First, Delbaen (2002) shows that $\mathrm{CVaR}_\delta(Z)$ is the smallest upper bound of $\mathrm{VaR}_\delta(Z)$ among all coherent risk measures that depend only on the distribution of Z. By relaxing the original problem to convex programs, a wide range of convex optimization algorithms can be employed to solve them. Second, optimization problems similar to model (32) have been investigated in the literature in the context of risk averse multistage stochastic programming (see, e.g., Guigues and Römisch 2012, and Philpott and de Matos 2012), thus similar algorithms and techniques can be employed. In addition, since model (32) and (33) take into account the extent of losses in the most adverse scenarios, they are also meaningful in their own right. In the following, we present a general regime to solve problem (32) using SAA approach. Since the discussion of model (33) resembles that of model (32) and is simpler, we will only discuss model (32).

SAA is a popular method in stochastic programming and has been used to deal with CVaR optimization problems, see Shapiro (2006) and Wang and Ahmed (2008). The concept of SAA is to approximate an expectation by the average of many samples generated by Monte Carlo method or other schemes. In problem (32), we have

$$\mathbb{E}[(\mathrm{CVaR}_\epsilon(H(\mathbf{x},\boldsymbol{\xi})) - \alpha)^+] \approx \frac{1}{N}\sum_{i=1}^{N}(\mathrm{CVaR}_\epsilon(H(\mathbf{x},\boldsymbol{\xi}_{F_i})) - \alpha)^+,$$

where N i.i.d. samples of Θ: $\Theta_1, \ldots, \Theta_N$ are generated from the distribution of Θ and the parameter of the distribution of $\boldsymbol{\xi}_{F_i}$ is Θ_i. Then, the original optimization problem can be approximated by the following problem:

$$\begin{aligned} \min_{\alpha\in\mathbb{R},\mathbf{x}\in\mathbb{R}^n,\mathbf{u}\in\mathbb{R}^N} \quad & \alpha + \frac{1}{(1-\delta)N}\mathbf{e}^T\mathbf{u} \\ \text{s.t.} \quad & u_i \geqslant \text{CVaR}_\epsilon(H(\mathbf{x},\boldsymbol{\xi}_{F_i})) - \alpha, \quad i = 1,2,...,N \\ & u_i \geqslant 0, \quad i = 1,2,...,N, \end{aligned} \tag{34}$$

where $\mathbf{e}$ denotes a vector of all ones. Notice that problem (34) is a CVaR constrained problem, and the CVaR in the constraint can also be approximated using SAA, resulting in the following problem:

$$\begin{aligned} \min_{\alpha,\mathbf{x},\mathbf{u},\mathbf{v}} \quad & \alpha + \frac{1}{(1-\delta)N}\mathbf{e}^T\mathbf{u} \\ \text{s.t.} \quad & u_i \geqslant v_i + \frac{1}{(1-\epsilon)M}\sum_{j=1}^{M}[H(\mathbf{x},\boldsymbol{\xi}_{F_i}^j) - v_i]^+ - \alpha, \quad i = 1,2,...,N \\ & u_i \geqslant 0, \quad i = 1,2,...,N, \end{aligned} \tag{35}$$

where $\mathbf{v} = (v_1, \ldots, v_N), i = 1, \ldots, N$ are auxiliary variables and $(\boldsymbol{\xi}_{F_i}^1, \ldots, \boldsymbol{\xi}_{F_i}^M), i = 1, \ldots, N$ are i.i.d. samples generated from F_i. When $H(\mathbf{x},\boldsymbol{\xi})$ is convex in $\mathbf{x}$, problem (35) is a convex program with a linear objective function and can be solved efficiently for large sized problems. Now we turn to the number of samples, N and M. For M, Wang and Ahmed (2008) shows that under certain regularity conditions, there is at least $1-\epsilon$ probability that the SAA of CVaR lies within the γ-neighborhood of CVaR when

$$M \geqslant \frac{C_1(H,F)}{\gamma^2}\left[C_2(H,F)n + C_3(H,F)\log\left(\frac{1}{\epsilon}\right)\right], \tag{36}$$

where $C_1(H,F)$, $C_2(H,F)$ and $C_3(H,F)$ are constants for a given objective function $H(\mathbf{x},\boldsymbol{\xi})$ and distribution F, n is the dimension of $\mathbf{x}$. Since the constants are typically difficult to calculate, the bound in (36) serves as a benchmark to estimate the order of M. For N, Shapiro (2006) proves that under mild conditions, the sample size

$$N \geqslant \frac{D_1(H,\mathbb{P}_1)}{\gamma^2}\left[n\log\left(\frac{D_2(H,\mathbb{P}_1)}{\gamma}\right) + \log\left(\frac{D_3(H,\mathbb{P}_1)}{\epsilon}\right)\right] \tag{37}$$

ensures that the solution of problem (34) lies within the γ-neighborhood of the solution of problem (32) with probability $1-\epsilon$. Here $D_1(H,\mathbb{P}_1)$, $D_2(H,\mathbb{P}_1)$ and $D_3(H,\mathbb{P}_1)$ are constants for a given objective function $H(\mathbf{x},\boldsymbol{\xi})$ and distribution $\mathbb{P}_1$ (of F), n is the dimension of $\mathbf{x}$.

Using the estimates in (36) and (37), we can see that the size of problem (35) is quite large. In practice, there are many ways to accelerate the computation. For instance, Alexander et al. (2006) and Xu and Zhang (2009) show that smoothing the CVaR can enhance the computation efficiency up to several times.

Remark 3. When the expression inside CVaR can be evaluated and differentiated easily, we only need to sample from distribution $\mathbb{P}_1$ and the resulting problem is much easier to solve. For instance, consider the following models:

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathrm{CVaR}_\delta(\mathrm{VaR}_\epsilon(\boldsymbol{\xi}^T\mathbf{x})), \tag{38}$$

$$\min_{\mathbf{x}\in\mathcal{X}} \ \mathrm{CVaR}_\delta(\mathrm{CVaR}_\epsilon(\boldsymbol{\xi}^T\mathbf{x})), \tag{39}$$

where the distribution of $\boldsymbol{\xi}$ is normal. Since $\mathrm{VaR}_\epsilon(\boldsymbol{\xi}^T\mathbf{x})$ and $\mathrm{CVaR}_\epsilon(\boldsymbol{\xi}^T\mathbf{x})$ have closed-form expressions (given in (28), (29)), problem (38) and (39) can be solved by second order cone programs (SOCP). For CVaR-Expectation problems, when the distribution of $\boldsymbol{\xi}$ is discrete, problem (33) reduces to a mean-CVaR problem and can be solved efficiently (Iyengar and Ma 2013 and Wen et al. 2013).

Having filled in several empty entries in Table 1, we have Table 2. In Table 2, the symbol $\sim$ means that the model is equivalent to another model labeled by the number and the symbol $\times$ means that such combination of risk measures has not been considered yet. (In Table 2, Var-Worst-Case and CVaR-Worst-Case model can be viewed as VaR-CVaR and CVaR-CVaR models, respectively, with the δ in the inner CVaR being 1.)

5. Numerical Experiments

In this section, we conduct numerical experiments to demonstrate the tractability and effectiveness of the models proposed in Section 4. In particular, we consider the VaR-Expectation model, the

Table 2 Different composites of risk measures: a full table

$\mu(\cdot)$ \ $g_F(\cdot)$	Expectation	CVaR	VaR	Worst-Case
Singleton	(1)	(4)	(3)	(5)
Expectation	$\sim$ (1)	$\times$	$\times$	$\times$
VaR	(21)	(31)	(30)	$\sim$(31)
Worst-Case	(6)	(12)	(11)	$\sim$(5)
CVaR	(33)	(32)	(38)	$\sim$(32)

CVaR-Expectation model and the CVaR-CVaR model and conduct numerical experiments using the portfolio selection problem. In the portfolio selection problem, a decision maker has to choose a portfolio to invest using available stocks based on historical data. In particular, our data set contains the daily returns of 359 different stocks that are in the S&P 500 index and do not have missing data from 2010 to 2011. In the following, we first show that when using a VaR-Expectation model, the corresponding distribution sets (for the VaR) depend on the decision, which is a distinguishing feature of the model. We show that this feature makes the VaR-Expectation model less conservative than the distributionally robust model in Delage and Ye (2010). Then we demonstrate that all the three models can be computed efficiently even when the sample size is large enough to ensure the precision of the solution. Finally, we compare the resulting returns of the three models with existing models.

5.1. Comparing the VaR-Expectation model and the distributionally robust model

We have discussed in Section 4 that one important feature of our proposed approach, the VaR-Expectation model, is that the distribution set (for the VaR) depends on the decision $\mathbf{x}$ and therefore the obtained solution from this model will be less conservative than that of a traditional distributionally robust model in which the distribution set is independent of $\mathbf{x}$. Here we illustrate this feature using numerical examples. Suppose a decision maker builds a portfolio of n stocks using the VaR-Expectation model. The joint distribution of the stocks is parameterized by an n-dimensional normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Σ. At each decision time, we

use the returns in the past t days to derive the posterior distribution for $\boldsymbol{\mu}$ and Σ. Let $\boldsymbol{\mu}_0$ and Σ_0 denote the mean and covariance of the empirical distribution of the stocks in the past t days. Using multivariate Jeffereys prior density as the prior distribution (which is the uninformative prior for normal distributions), we have that the posterior distribution $\mathbb{P}_1$ of $\boldsymbol{\mu}$ and Σ is given by (Gelman et al. 2013):

$$f(\boldsymbol{\mu},\Sigma) = \mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, t^{-1}\Sigma)\mathcal{W}^{-1}(\Sigma|t\Sigma_0, t-1), \tag{40}$$

where $\mathcal{N}(\boldsymbol{\mu}|\boldsymbol{\mu}_0, t^{-1}\Sigma)$ is the probability density function of a multivariate normal distribution with mean $\boldsymbol{\mu}_0$ and covariance matrix $t^{-1}\Sigma$ and $\mathcal{W}^{-1}(\Sigma|t\Sigma_0, t-1)$ is the probability density function of an inverse-Wishart distribution with scale matrix $t\Sigma_0$ and degree of freedom $t-1$. In the following, we set $t = 30$. For each fixed decision (allocation) $\mathbf{x}$, the worst-case distribution $F_0(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_0(\mathbf{x}), \Sigma_0(\mathbf{x}))$ of loss $\boldsymbol{\xi}$ satisfies the following condition:

$$\mathbb{P}_1(\mathbb{E}_F[\boldsymbol{\xi}^T\mathbf{x}] \geqslant \mathbb{E}_{F_0(\mathbf{x})}[\boldsymbol{\xi}^T\mathbf{x}]) = 1-\delta,$$

or equivalently,

$$\mathbb{P}_1(\boldsymbol{\mu}_F^T\mathbf{x} \geqslant \boldsymbol{\mu}_0(\mathbf{x})^T\mathbf{x}) = 1-\delta.$$

Thus, the distribution set of the VaR-Expectation model is:

$$\mathcal{F}_{\mathbf{x}} = \left\{\mathcal{N}(\boldsymbol{\mu},\Sigma)|\boldsymbol{\mu}^T\mathbf{x} \leqslant \boldsymbol{\mu}_0(\mathbf{x})^T\mathbf{x}\right\},$$

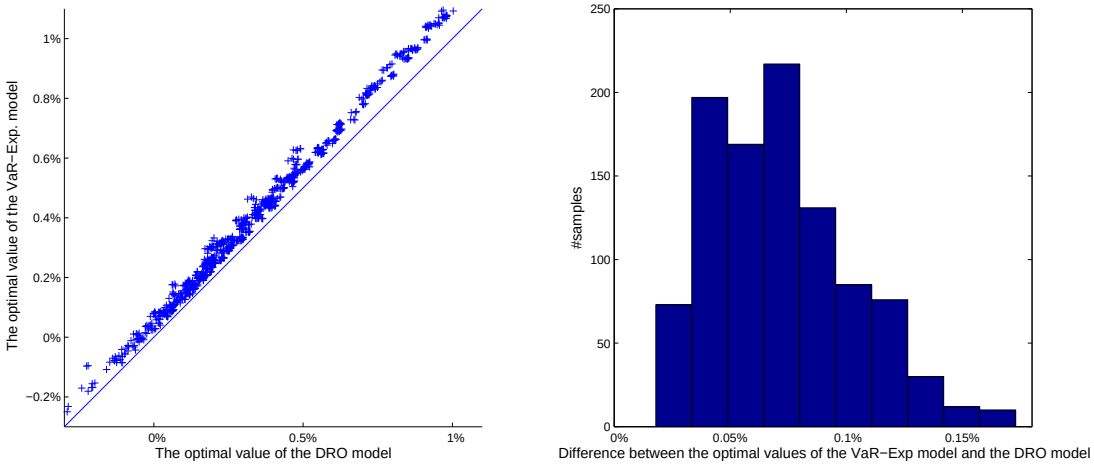
which depends on $\mathbf{x}$. The optimization problem under the VaR-Expectation model is (we assume that the total investment amount must equal to 1):

$$\max_{\mathbf{x}\geqslant 0, \mathbf{e}^T\mathbf{x}=1} \boldsymbol{\mu}_0(\mathbf{x})^T\mathbf{x}. \tag{41}$$

To compare to the distributionally robust model in Delage and Ye (2010), we note that the objective function in Delage and Ye (2010) is

$$\max_{\mathbf{x}\geqslant 0, \mathbf{e}^T\mathbf{x}=1} \min_{F\in\mathcal{F}} \boldsymbol{\mu}_F^T\mathbf{x} \tag{42}$$

Figure 1 Comparison between the VaR-Expectation model and the DRO model.

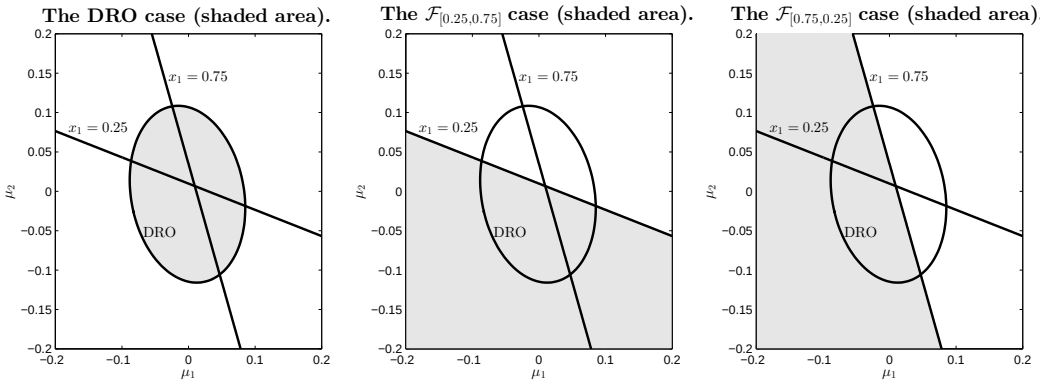


where

$$\mathcal{F} = \left\{ F(\boldsymbol{\xi}) \in \mathcal{M} \left| \begin{array}{l} (\mathbb{E}_{\boldsymbol{\xi}\sim F}(\boldsymbol{\xi}) - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\mathbb{E}_{\boldsymbol{\xi}\sim F}(\boldsymbol{\xi}) - \boldsymbol{\mu}_0) \leqslant \gamma_1 \\ \\ \mathbb{E}_{\boldsymbol{\xi}\sim F}[(\boldsymbol{\xi} - \boldsymbol{\mu}_0)(\boldsymbol{\xi} - \boldsymbol{\mu}_0)^T] \preceq \gamma_2 \Sigma_0 \end{array} \right. \right\},$$

where γ_1 and γ_2 are chosen according to the discussions in Delage and Ye (2010) to ensure that with probability δ, $\mathcal{F}$ contains the true distribution. In the following, we fix $\delta = 0.95$. We perform the following procedures: we randomly pick two stocks and use their empirical returns from 3/4/2010 to 4/14/2010 to fit a normal distribution. Then we draw 10^6 data from the distribution to form the data set. (In Delage and Ye 2010, it usually needs at least 10^5 data points to get a valid γ_1 and γ_2, therefore we have to take this bootstrapping method.) Then we solve (41) and (42) respectively using these data. By the discussions in Section 4.1, the optimal value of (41) should be larger than that of (42). In our numerical experiment, we repeat the above procedures 1000 times and plot the result in Figure 1. The left figure in Figure 1 shows a scatter plot of the optimal values obtained by the distributionally robust model in Delage and Ye (2010) (x-axis) and the VaR-Expectation model (y-axis) in the 1000 experiments, and the right figure shows the distribution of the difference of the optimal values in the same set of experiments. From Figure 1, we can see that the VaR-Expectation model results in higher value in all the cases, meaning that it is less

Figure 2 Comparison of distribution sets.



conservative. Indeed, the average optimal value of (41) in the 1000 experiments is 0.070% larger than that of (42). Moreover, for one particular experiment, we plot the projection of $\mathcal{F}_{[0.75,0.25]}$, $\mathcal{F}_{[0.25,0.75]}$ and the distribution set of the DRO model on the plane spanned by $\boldsymbol{\mu}$ in Figure 2. From Figure 2, we can see that in the DRO approach, the distribution set is independent of the choice of $\mathbf{x}$. However, in the VaR-Expectation model, the distribution set changes with the choice of $\mathbf{x}$. As we have mentioned above, it is such feature of the VaR-Expectation model that makes the solution less conservative yet with the same level of probabilistic guarantee.

5.2. Solving the composite risk measure model

Under the setting in Section 5.1, the VaR-Expectation model can be solved by the alternating direction augmented Lagrangian method (ADM, see Wen et al. 2013), the CVaR-Expectation model can be solved by an LP, and the CVaR-CVaR model can be solved by an SOCP. We evaluate the performance of these three models with different stock number n and sample size N. Our experiments are performed on a laptop with 8.00 GB of RAM and 2.20 GHz processor, using MOSEK with the Matlab interface.

We first solve the VaR-Expectation, CVaR-Expectation and CVaR-CVaR models for different sample size N when $t = 30$ and $n = 4$, while both outer and inner risk levels are chosen as 0.95. The stocks are randomly chosen from all 359 stocks and the period we consider is from 3/4/2010 to 4/14/2010. We perform the experiment on the same set of stocks 100 times, and the results are

Table 3 Computation time and precision of different CRM models when $n = 4$.

#samples	VaR-Expectation			CVaR-Expectation			CVaR-CVaR		
N	ave	std	time	ave	std	time	ave	std	time
5000	0.0005	1.91×10^{-4}	5.12	-0.0007	1.44×10^{-4}	0.13	-0.0370	2.80×10^{-4}	0.76
10000	0.0004	1.56×10^{-4}	5.17	-0.0007	7.94×10^{-5}	0.16	-0.0371	2.01×10^{-4}	1.78
20000	0.0004	1.09×10^{-4}	4.99	-0.0007	6.68×10^{-5}	0.25	-0.0371	1.44×10^{-4}	3.26
50000	0.0005	6.43×10^{-5}	7.58	-0.0007	4.20×10^{-5}	0.70	-0.0372	7.94×10^{-5}	10.02
100000	0.0004	4.63×10^{-5}	9.26	-0.0007	3.18×10^{-5}	1.65	-0.0373	7.17×10^{-5}	17.64

shown in Table 3 and Figure 3. In Table 3, the first column, denoted by "ave", for each method is the average of the optimal values of the corresponding models, the second column, denoted by "std", is the standard deviation of the optimal values in all experiments, while the third column is the average computation time (in seconds). Notice that when $N = 100000$, all these three models can still be computed efficiently. In the mean time, the solutions of SAA problems can approximate the solution of the original models very well. Figure 3 displays the scatter plots of the weights of the first 3 stocks in the optimal portfolio of all experiments (the weight of the last stock can be computed by one minus the total weights of the first three stocks). The result shows that the larger N is, the more concentrated the solution is. It indicates that the optimal solution also converges as the sample size becomes large.

Now we fix $N = 5000$ and choose different n, while the period considered and all other parameters are the same as in the previous experiment. The results are presented in Table 4, where t_{ave} is the average computation time in all experiments, t_{min} is the minimum computation time and t_{max} is the maximum computation time. The models considered are (from left to right in Table 4): VaR-Expectation model, CVaR-Expectation model and CVaR-CVaR model. We can see that our new models can still be computed in a reasonably amount of time even when n is large.

5.3. Comparing CRM models with existing models using real market data

To test the performance of our new models in real world trading operations, we compare our models with existing models for portfolio selection problem. In each experiment, we randomly

Figure 3 Optimal allocations of the approximated problems.

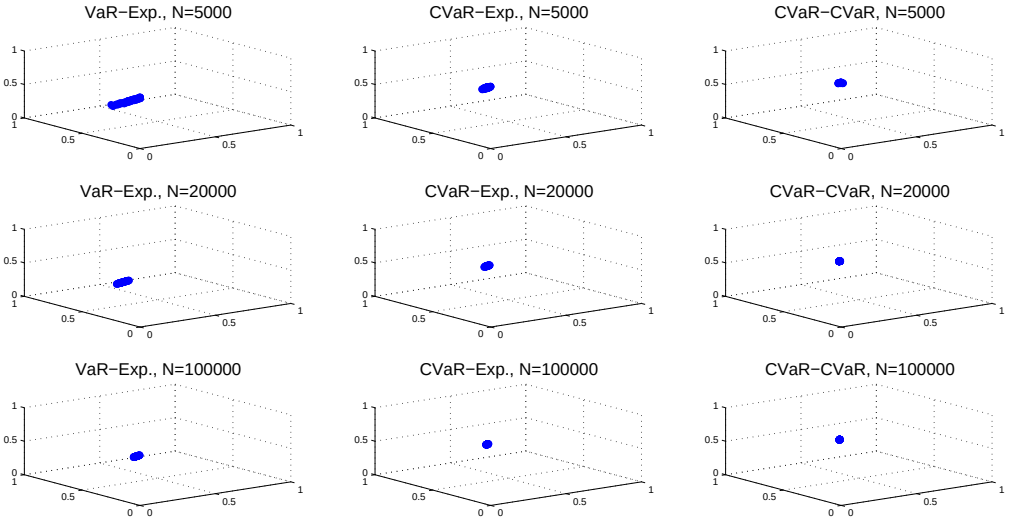


Table 4 Computation time of different CRM models when $N = 5000$.

#stocks	VaR-Expectation			CVaR-Expectation			CVaR-CVaR		
n	t_{ave}	t_{min}	t_{max}	t_{ave}	t_{min}	t_{max}	t_{ave}	t_{min}	t_{max}
10	4.52	3.93	4.97	0.14	0.13	0.20	3.17	2.71	5.72
20	4.49	3.98	5.07	0.18	0.17	0.22	6.77	5.22	12.48
30	5.36	4.57	6.16	0.22	0.20	0.25	12.62	10.39	19.43
40	6.19	5.44	7.07	0.28	0.25	0.32	19.49	15.40	29.15
50	6.38	5.60	7.11	0.33	0.29	0.39	29.33	22.33	41.29

choose 4 stocks from all 359 stocks to build a dynamic portfolio during the period from 3/4/2010 to 4/27/2011 (300 days in total). The portfolio is recomputed everyday using the returns of the last 30 days as input data. At each day of the experiment, only the returns of the last 30 days can be used. We set sample size as 2000 for VaR-Expectation, CVaR-Expectation and CVaR-CVaR models, and compare the results with distributionally robust model (Delage and Ye 2010), worst-case VaR model (El Ghaoui et al. 2003) and single stock (SS) model. The single stock model chooses the stock that has the highest average return rate in the last 30 days as the sole stock for that day, and is used as a naive benchmark here. The average cumulative return of each day is shown in Figure 4, while the average standard deviation and daily return are presented in Table

Figure 4 Comparison of wealth accumulation of all models.

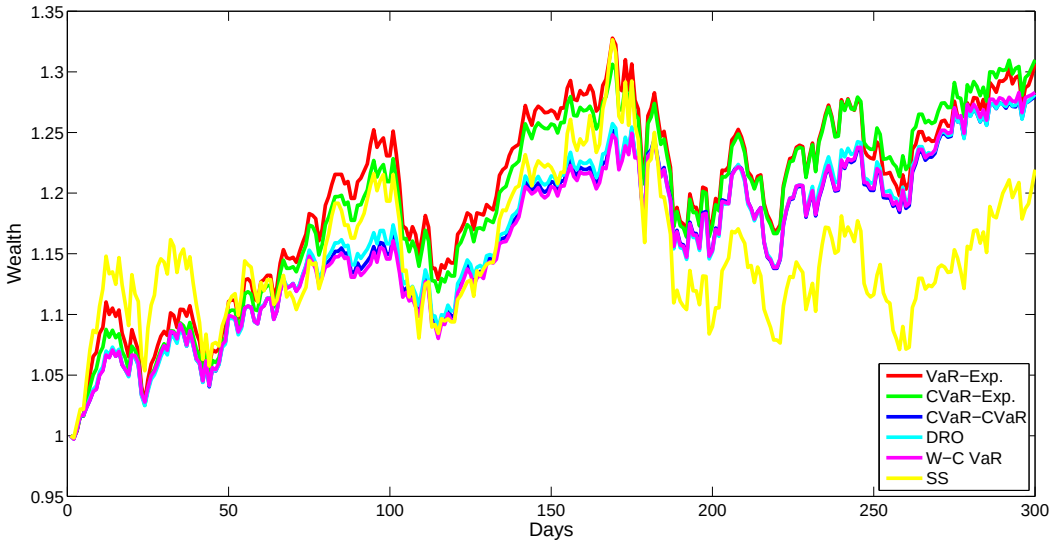


Table 5 Average return rate and standard deviation of all models.

	VaR-Exp.	CVaR-Exp.	CVaR-CVaR	DRO	W-C VaR	SS
ave return	0.096 %	0.096 %	0.087 %	0.088 %	0.087 %	0.078 %
ave std	1.49×10^{-2}	1.34×10^{-2}	1.17×10^{-2}	1.19×10^{-2}	1.17×10^{-2}	2.17×10^{-2}

5. In this experiment, the volatilities of our models are significantly smaller than that of the SS approach. The CVaR-Expectation model and VaR-Expectation model achieve better return rates than the other models, while CVaR-CVaR model performs as good as DRO model and worst-case VaR model. Though we cannot draw general conclusions about which model is intrinsically better without more intensive tests, this experiment shows that the CRM models choose portfolios with robust performance, which is a property we desire.

6. Conclusions

In this paper, we propose a unified framework for decision making under uncertainty using the composite of risk measures. Our focus has been the case where the distribution of uncertain model parameters can be parameterized by a finite number of parameters, which includes a large family of problems. The generality of our framework allows us to unify several existing models as well as to construct new models within the framework. We show through theoretical proofs and numerical

experiments that our new paradigms yield less conservative solutions, yet provide the same degree of probabilistic guarantee.

References

Alexander, S., T. F. Coleman, Y. Li. 2006. Minimizing CVaR and VaR for a portfolio of derivatives. *Journal of Banking & Finance* **30**(2) 583–605.

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

Ben-Tal, A., D. den Hertog, A. De Waegenaere, B. Melenberg, G. Rennen. 2013. Robust solutions of optimization problems affected by uncertain probabilities. *Management Science* **59**(2) 341–357.

Ben-Tal, A., L. El Ghaoui, A. Nemirovski. 2009. *Robust Optimization*. Princeton University Press.

Ben-Tal, A., A. Nemirovski. 1999. Robust solutions of uncertain linear programs. *Operations Research Letters* **25**(1) 1–13.

Bertsimas, D., D. B. Brown. 2009. Constructing uncertainty sets for robust linear optimization. *Operations Research* **57**(6) 1483–1495.

Bertsimas, D., D. B. Brown, C. Caramanis. 2011. Theory and applications of robust optimization. *SIAM Review* **53**(3) 464–501.

Bertsimas, D., V. Gupta, N. Kallus. 2014. Data-driven robust optimization. Working paper.

Bertsimas, D., D. Pachamanova, M. Sim. 2004. Robust linear optimization under general norms. *Operations Research Letters* **32**(6) 510–516.

Bertsimas, D., I. Popescu. 2005. Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* **15**(3) 780–804.

Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Birge, J. R., F. Louveaux. 2011. *Introduction to Stochastic Programming*. Springer.

Chen, X., M. Sim, P. Sun. 2007. A robust optimization perspective on stochastic programming. *Operations Research* **55**(6) 1058–1071.

Cui, X., S. Zhu, X. Sun, D. Li. 2013. Nonlinear portfolio selection using approximate parametric value-at-risk. *Journal of Banking & Finance* **37**(6) 2124–2139.

Dantzig, G. B. 1955. Linear programming under uncertainty. *Management Science* **1**(3-4) 197–206.

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

Delbaen, F. 2002. Coherent risk measures on general probability spaces. K. Sandmann, P. J. Schoenbucher, eds., *Advances in Finance and Stochastics*. Springer Berlin Heidelberg, 1–37.

Dupačová, J. 1987. The minimax approach to stochastic programming and an illustrative application. *Stochastics: An International Journal of Probability and Stochastic Processes* **20**(1) 73–88.

Einhorn, D., A. Brown. 2008. Private profits and socialized risk. *Global Association of Risk Professionals* **42** 10–26.

El Ghaoui, L., H. Lebret. 1997. Robust solutions to least-squares problems with uncertain data. *SIAM Journal on Matrix Analysis and Applications* **18**(4) 1035–1064.

El Ghaoui, L., M. Oks, F. Oustry. 2003. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Operations Research* **51**(4) 543–556.

El Ghaoui, L., F. Oustry, H. Lebret. 1998. Robust solutions to uncertain semidefinite programs. *SIAM Journal on Optimization* **9**(1) 33–52.

Föllmer, H., A. Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6**(4) 429–447.

Gaivoronski, A. A., G. Pflug. 2005. Value-at-risk in portfolio optimization: properties and computational approach. *Journal of Risk* **7**(2) 1–31.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin. 2013. *Bayesian Data Analysis*. CRC press.

Guigues, V., W. Römisch. 2012. Sampling-based decomposition methods for multistage stochastic programs based on extended polyhedral risk measures. *SIAM Journal on Optimization* **22**(2) 286–312.

Huber, P. J. 1981. *Robust Statistics*. Wiley series in probability and mathematical statistics, Wiley, New York.

Iyengar, G., A. K. C. Ma. 2013. Fast gradient descent method for mean-CVaR optimization. *Annals of Operations Research* **205**(1) 203–212.

Kast, R., E. Luciano, L. Peccati. 1998. VaR and optimization. *2nd International Workshop on Preferences and Decisions, Trento, July*, vol. 1. 1998.

Klabjan, D., D. Simchi-Levi, M. Song. 2013. Robust stochastic lot-sizing by means of histograms. *Production and Operations Management* **22**(3) 691–710.

Lucas, A., P. Klaassen. 1998. Extreme returns, downside risk, and optimal asset allocation. *The Journal of Portfolio Management* **25**(1) 71–79.

Luedtke, J., S. Ahmed. 2008. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal on Optimization* **19**(2) 674–699.

Natarajan, K., D. Pachamanova, M. Sim. 2009. Constructing risk measures from uncertainty sets. *Operations Research* **57**(5) 1129–1141.

Nemirovski, A., A. Shapiro. 2006. Scenario approximations of chance constraints. *Probabilistic and randomized methods for design under uncertainty*. Springer, 3–47.

Pardo, L. 2005. *Statistical Inference Based on Divergence Measures*. CRC Press.

Philpott, A. B., V. L. de Matos. 2012. Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion. *European Journal of Operational Research* **218**(2) 470–483.

Prékopa, A. 1995. *Stochastic Programming*. Springer.

RiskMetrics. 1996. *Riskmetrics: Technical Document*. Morgan Guaranty Trust Company of New York.

Rockafellar, R. T., S. Uryasev. 2000. Optimization of conditional value-at-risk. *Journal of Risk* **2** 21–42.

Sarykalin, S., G. Serraino, S. Uryasev. 2008. Value-at-risk vs. conditional value-at-risk in risk management and optimization. *Tutorials in Operations Research. INFORMS, Hanover, MD* .

Scarf, H., K. J. Arrow, S. Karlin. 1958. A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* **10** 201–209.

Shapiro, A. 2006. On complexity of multistage stochastic programs. *Operations Research Letters* **34**(1) 1–8.

Shapiro, A., D. Dentcheva, A. Ruszczyński. 2009. *Lectures on Stochastic Programming: Modeling and Theory*, vol. 9. SIAM.

Soyster, A. L. 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research* **21**(5) 1154–1157.

Wang, W., S. Ahmed. 2008. Sample average approximation of expected value constrained stochastic programs. *Operations Research Letters* **36**(5) 515–519.

Wang, Z., P. Glynn, Y. Ye. 2013. Likelihood robust optimization for data-driven problems. Working paper.

Wen, Z., X. Peng, X. Liu, X. Sun, X. Bai. 2013. Asset allocation under the basel accord risk measures. Working paper.

Wiesemann, W., D. Kuhn, M. Sim. 2014. Distributionally robust convex optimization. *Operations Research* .

Xu, H., D. Zhang. 2009. Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. *Mathematical Programming* **119**(2) 371–401.

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