manuscript

Target Oriented Distributionally Robust Optimization and Its Applications to Surgery Allocation

Vincent Tsz Fai Chow, Zheng Cui, Daniel Zhuoyu Long
Department of Systems Engineering and Engineering Management, Chinese University of Hong Kong, Hong Kong, China
tfchow@se.cuhk.edu.hk, zcui@se.cuhk.edu.hk zylong@se.cuhk.edu.hk

In this paper, we propose a decision criterion that characterizes an enveloping bound on monetary risk measures and is computationally friendly. We start by extending the classical Value at Risk (VaR) measure. While VaR evaluates the threshold loss value such that the loss from the risk position exceeding that threshold is at a given probability level, it fails to indicate a performance guarantee at other probability levels. We define the Probabilistic Enveloping Measure (PEM) to establish the bound information for the tail probability of the loss at all levels. Using a set of normative properties, we then generalize the PEM to the Risk Enveloping Measure (REM) such that the bound on the general monetary risk measures at all levels of risk aversion are captured. The coherent version of the REM (CREM) is also investigated. We demonstrate its applicability by showing how the coherent REM can be incorporated in distributionally robust optimization. Specifically, we apply the CREM criterion in the surgery block allocation problems and provide a formulation that can be efficiently solved. Based on this application, we report favorable computational results from optimizing over the CREM criterion.

Key words: risk measure; enveloping bound; distributionally robust optimization; surgery block allocation
History:

---

1. Introduction

When the outcomes in a decision making problem are uncertain, we typically evaluate the outcome by its expectation and optimize over it. This happens no matter which subject area we consider — we minimize the expected cost in inventory control problems, maximize the expected return in portfolio optimization problems, and minimize the expected waiting times in a hospital when determining the optimal staffing requirement (Ahmed and Alkhamis 2009). In the past few decades, however, an increasing number of researchers have become aware of that solely focusing on the expected outcome fails to take into account the associated risk. To incorporate the practical concern of risk, numerous studies have adopted the expected utility as a substitute for the expected outcome to evaluate the uncertain performance. Examples can be seen in the application in the newsvendor problem (Eeckhoudt et al. 1995), the multi-period inventory control problem (Chen et al. 2007), and the reserve capacity determination for a hospital in response to uncertain demand (Rodriguez-Alvarez et al. 2012). In addition to using the expected utility, valuating the uncertain outcome by the monetary risk measure can provide a clear quantitative indication on the associated risk. It has also been applied to the newsvendor problems (Choi and Ruszczyński 2008, Choi et al. 2011), the inventory-pricing problem (Chen et al. 2009), and the hospital staffing problems (Pender 2016). We begin with the second approach, the monetary risk measure, for its intuitive implication on the riskiness. Specifically, we aim to provide a comprehensive description of riskiness by establishing an enveloping bound on monetary risk measures. We then demonstrate its connection with the first approach, the expected utility theory.

Among monetary risk measures, Value at Risk (VaR) is a ubiquitous one that has wide applications from finance to supply chain management. An extensive overview of the applicability of VaR can be seen in Jorion (2006). Given a probability level, α, the VaR of a risk position, $\tilde{x}$, is defined as a threshold loss value, such that the probability that the loss on the position exceeding this threshold is no more than $(1-\alpha)$. Mathematically,

$$\mathrm{VaR}_{\alpha}(\tilde{x}) = \inf\left\{t \;:\; \mathbb{P}(-\tilde{x} > t) \leq 1-\alpha\right\}, \tag{1}$$

where $\mathbb{P}$ denotes the probability measure. Hence, VaR is a quantile-based monetary risk measure that quantifies the risk of the position $\tilde{x}$ via the α-quantile of the loss, $-\tilde{x}$.

As an illustrative example, Table 1 lists three risk positions ($\tilde{x}_1, \tilde{x}_2$ and $\tilde{x}_3$), among which the investor will select exactly one. By varying the probability level α, the VaR in Table 1 provides a snapshot of the risk distributions at various quantiles of interest. Specifically, for Position $\tilde{x}_1$, $\mathrm{VaR}_{0.9}(\tilde{x}_1) = 500$, i.e., $\mathbb{P}(-\tilde{x}_1 > 500) \leq 1-0.9$; it implies the probability that $\tilde{x}_1$ results in a loss exceeding \$500 is less than 10%. By contrast, the investor might have a 10% chance to incur a \$2,000 loss if Position $\tilde{x}_2$ is chosen. Thus, given the probability level $\alpha = 0.9$, the investor prefers Position $\tilde{x}_1$ over $\tilde{x}_2$. However, for a decision criterion, the evaluation of risk that is based on a single quantile parameter α may not be adequate. Position $\tilde{x}_1$ is the most preferred position at $\alpha \in \{0.9, 0.95\}$, while $\tilde{x}_2$ is most preferred at $\alpha = 0.98$. While Position $\tilde{x}_1$ is preferred at lower quantiles, the associated loss at $\alpha = 0.98$ is significantly higher than those in other positions. It thus has a low probability of a

Positions	$\alpha = 0.9$	$\alpha = 0.95$	$\alpha = 0.98$
$\tilde{x}_1$	500	1,000	20,000
$\tilde{x}_2$	2,000	2,000	2,000
$\tilde{x}_3$	1,000	1,800	2,100

Table 1 VaR_α **of three risk positions**

potentially catastrophic level of losses. On the other hand, Position $\tilde{x}_2$ has the highest VaR at lower quantiles but the lowest risk at $\alpha = 0.98$. However, when considering the statistics of the loss quantiles at various levels collectively, Position $\tilde{x}_3$ may be preferred over the other two. Therefore, while favoring positions with low quantiles at a specific probability level, VaR has no performance guarantee at any other probability level. In many practical problems, it has been well documented that decision makers are unavoidably sensitive to performance at more than one probability level (Payne et al. 1980, 1981).

Motivated by the concept of VaR and the need to encompass a range of quantile values in evaluating risk positions, we propose a class of decision criteria that allows us to specify the probabilistic bounds of losses exceeding thresholds at all levels. However, in general, the optimization problems involving VaR are already computationally difficult to solve (Nemirovski 2012). Therefore, a key issue in taking into account probabilistic bounds at a range of levels is not placing additional burden on the computational complexity.

Apart from VaR, other monetary risk measures have also received a great deal of attentions in the past decade, such as Conditional Value at Risk (CVaR) and Entropic Risk Measure. Interested readers can refer to Föllmer et al. (2004) for more details. Similar to the probability level α in VaR, the other monetary risk measures are related to parameters that characterize the level of risk aversion. For example, while using CVaR, a corresponding probability level must be declared; and if using Entropic Risk Measure, a risk aversion parameter has to be specified as an indicator of the level of concavity for the underlying exponential function curve. It is also of great interest to investigate how to incorporate the enveloping bound on a general monetary risk measure at all levels of risk aversion. Take the CVaR as an example. Given a particular probability level α, the monetary risk measure CVaR_α evaluates the expected loss beyond the α-quantile. However, it gives no information on the expected loss beyond the quantile at any other probability level. In this paper, we are going to propose a systematic index that can provide information on the bound of CVaR at all levels of α.

The concept of the enveloping bound can be traced back to the 1970s in the area of signal processing (e.g., Evans et al. 1977a,b). Xu et al. (2012) investigate the optimization problem with the presence of enveloping chance constraints. In the above studies, the enveloping bound serves a constraint such that a feasible solution has to be subject to the infinite set of constraints. In contrast, our paper incorporates the enveloping bound in the objective function. Hence, by solving the corresponding optimization problem, we can design a solution with the lowest risk from an enveloping perspective. This paper also generalizes Xu et al. (2012) in the sense that the concept of the enveloping bound can be embedded with any monetary risk measure, rather than only focusing on the enveloping chance constraints in Xu et al. (2012).

Driven by the enveloping bound, our new optimization framework is then applied on a surgery blocks assignment problem, which is a healthcare application from the bin packing problems. In this problem, the decision maker selects a subset of operating theaters to open, and assigns surgery cases to an open operating theater. Given the uncertain duration of a surgery, the decision maker has to mitigate the overtime risk, or there will be operational issues when the assignment plan is put into practice — the resources get locked up while some patients are in urgent need. Denton et al. (2010) is the first attempt to apply the robust linear program with box uncertainty on this problems. After that, Shylo et al. (2013) use a chance-constraint model to bound the overtime risk. Deng and Shen (2016) use the same approach and formulate a two-stage mixed integer programs without the need to assume normality on surgery durations. Recently, Zhang et al. (2018) tackle the same problem by formulating a distributionally robust chance-constrained problem. In their ambiguity sets, the information of means and covariances is included. In this paper, we use our approach on this problem and compare the computational performance with those from the benchmarks.

In summary, we develop a decision criterion that can both encompass the bound on quantiles at all probability levels and add no significant computational burden into the stochastic optimization problems. We then extend the idea to obtain a class of decision criteria that indicate an enveloping bound on any monetary risk measure for all levels of risk aversion. After that, we develop a solution approach to optimize the decision criteria in the decision making problem. Finally, we apply the framework on the surgery assignment problem and compare its performance against the existing method in the literatures.

Notations: We denote a random variable by a character with the tilde sign, such as $\tilde{x}$. We denote the probability space as $(\Omega, \mathcal{F}, \mathbb{P})$. An inequality between two risk positions such as $\tilde{x} \geq \tilde{y}$ denotes state-wise dominance, i.e., it is equivalent to $\tilde{x}(\omega) \geq \tilde{y}(\omega)$ for all $\omega \in \Omega$. In addition, a strict inequality such as $\tilde{x} > \tilde{y}$ implies that $\exists \epsilon > 0$ such that $\tilde{x} \geq \tilde{y} + \epsilon$. Finally, we define $\inf \emptyset = \infty$. All proofs appear in Appendix A in the online supplement.

2. Risk Enveloping Measure

In this section, we first extend VaR to incorporate a range of quantiles, and we then generalize it to a class of general risk measures called the *Risk Enveloping Measure*.

DEFINITION 1. Given a non-increasing function $\beta : \Re_+ \to [0, 1)$, we define the corresponding Probabilistic Enveloping Measure (PEM) as

$$\mathrm{PEM}_\beta(\tilde{x}) = \inf\{k \in \Re_+ \ : \ \mathbb{P}(-\tilde{x} > k\theta) \leq \beta(\theta), \ \forall \, \theta > 0\}. \tag{2}$$

Notice that here we use the term *enveloping* since we impose a bound on the tail probability at all levels, and correspondingly, we call the function $\beta(\theta)$ the *envelope function*. In particular, PEM is defined such that for all $\theta > 0$,

$$\mathbb{P}(-\tilde{x} > \mathrm{PEM}_\beta(\tilde{x})\theta) \leq \beta(\theta). \tag{3}$$

The envelope function $\beta(\theta)$ bounds the tail probability of the loss from the risk position $\tilde{x}$ exceeding the level $\mathrm{PEM}_\beta(\tilde{x})\theta$. When $\mathrm{PEM}_\beta(\tilde{x}) > 0$, we have $\mathbb{P}(-\tilde{x} > \theta) \leq \beta(\theta / \mathrm{PEM}_\beta(\tilde{x}))$. Therefore, a position with a low value of PEM_β is preferable because the probability that the loss exceeds any level θ is bounded by a low value. For example, by choosing an exponential envelope function $\beta(\theta) = \exp(-\theta)$, we illustrate the enveloping bound $\exp(-\theta/\mathrm{PEM})$ with different values of PEM in Figure 1.

Remark: Noticing that according to Equation (3), different choices of the envelope function $\beta(\theta)$ correspond to different shapes of the enveloping bound on the tail probability. Therefore, in practice, the decision-maker might choose the function $\beta(\theta)$ depending on the needs of bounding the tail probabilities. For example, if aiming at a more stringent restriction on the long tail probability, a β function which decays faster might be more suitable. Moreover, one might also take into account the computational tractability when choosing $\beta(\theta)$ since in different applications, different envelope functions lead to different level of computational complexity.

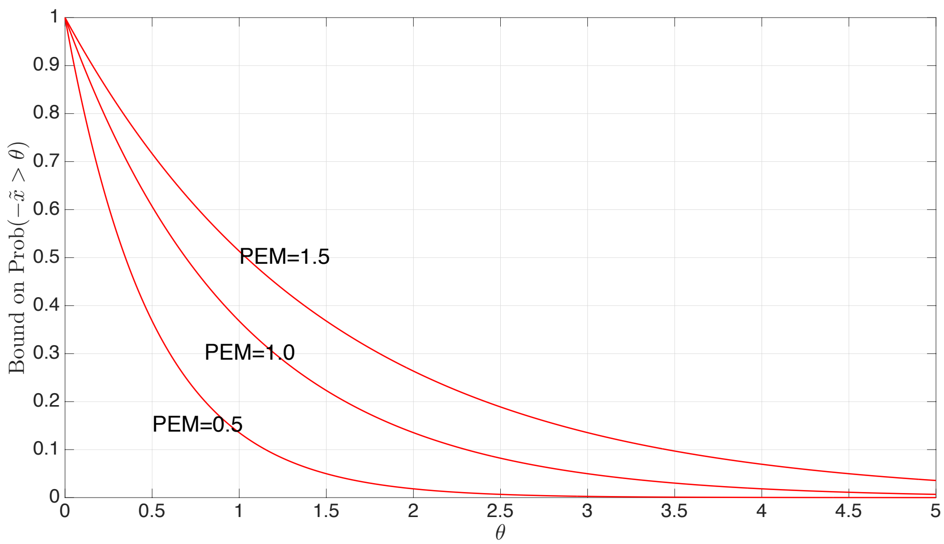

Figure 1 **Probabilistic bounds on the losses using an exponential envelope function.**

By enforcing constraints on the quantile at all probabilistic levels, the PEM essentially provides an enveloping bound on VaR, i.e., a bound on VaR at all probabilistic levels. Specifically, by the definition in Equations (1) and (2), we have

$$\mathrm{PEM}_{\beta}(\tilde{x}) = \inf\left\{k \in \Re_+ \ : \ \mathrm{VaR}_{1-\beta(\theta)}(\tilde{x}) \le k\theta, \ \forall\, \theta > 0\right\}. \tag{4}$$

Consider the case in which β^{-1}, the inverse function of β, is well defined on $[0,1)$. Then, Equation (4) explicitly implies the enveloping bound, $\mathrm{VaR}_{\alpha}(\tilde{x}) \le \mathrm{PEM}_{\beta}(\tilde{x}) \cdot \beta^{-1}(1-\alpha)\ \forall \alpha \in (0,1]$. Thus, in problems of optimization under uncertainty, using PEM as an objective function can lead to a solution with a performance guarantee on VaR at all probabilistic levels.

While PEM carries bounding information for VaR at all probabilistic levels, it is also of great interest for extending the enveloping bound to the general monetary risk measure. To this end, we define the Risk Enveloping Measure by generalizing the definition of PEM. We first note that VaR belongs to the class of monetary risk measures that have several salient properties. Let $\mathcal{X}$ be the set of all risk positions.

Definition 2. A function $\mu : \mathcal{X} \to \Re$ is a monetary risk measure if it satisfies the following properties for all $\tilde{x}, \tilde{y} \in \mathcal{X}$:
(P1) Monotonicity: if $\tilde{x} \ge \tilde{y}$, then $\mu(\tilde{x}) \le \mu(\tilde{y})$.
(P2) Positive Homogeneity: $\mu(k\tilde{x}) = k\mu(\tilde{x})$ for all $k \ge 0$.
(P3) Translation Invariance: for any constant position c, $\mu(\tilde{x} + c) = \mu(\tilde{x}) - c$.

Remark: Note that in the literature (e.g., Föllmer et al. 2004), monetary risk measure is defined solely by Monotonicity and Translation Invariance. Here we also impose Positive Homogeneity for two reasons. First, it is an important property of VaR (Jorion 2006). Secondly, we aim to specifically focus on the measures with this cardinal nature — that is, $2\tilde{x}$ is "twice" as risky as "$\tilde{x}$" rather than just simply riskier (Artzner et al. 1999). We observe that Positive Homogeneity implies $\mu(0) = 0$.

Originating from VaR, which belongs to the class of monetary risk measures, PEM satisfies the properties of Monotonicity and Positive Homogeneity. However, it violates the property of Translation Invariance. Instead, PEM has the property of Satisficing (proposed in Brown and Sim 2009). We explore the generalization of PEM by proposing a new class of criteria termed the *Risk Enveloping Measure* as follows:

Definition 3. A function $\rho : \mathcal{X} \to [0, \infty]$ is a Risk Enveloping Measure (REM) if it satisfies the following properties for all $\tilde{x}, \tilde{y} \in \mathcal{X}$:

(P1) Monotonicity: if $\tilde{x} \geq \tilde{y}$, then $\rho(\tilde{x}) \leq \rho(\tilde{y})$.

(P2) Positive Homogeneity: $\rho(k\tilde{x}) = k\rho(\tilde{x})$ for all $k \geq 0$.

(P3) Satisficing:

(a) if $\tilde{x} \geq 0$, then $\rho(\tilde{x}) = 0$;

(b) if $\tilde{x} < 0$, then $\rho(\tilde{x}) = \infty$.

(P4) Right Continuity: $\lim_{a \downarrow 0} \rho(\tilde{x} + a) = \rho(\tilde{x})$.

The properties of Monotonicity and Positive Homogeneity are inherited from monetary risk measures. The Satisficing property is essentially related to the two extreme cases of the enveloping bound. In particular, when $\tilde{x} \geq 0$, there is no risk at all for the whole class of monetary risk measures; hence, the associated REM should be zero. Take the PEM in Definition 1 as an example. For PEM, the constraint $\mathbb{P}(-\tilde{x} > k\theta) \leq \beta(\theta),\ \forall\, k, \theta > 0$ is always satisfied in the case of $\tilde{x} \geq 0$ and thus the PEM is zero. In contrast, if $\tilde{x} < 0$, the associated REM should be the maximum because there is always non-negligible risk regardless of the underlying monetary risk measure. The justification for Satisficing can also be seen in Brown and Sim (2009). The Right Continuity property implies that if an infinitesimally small but positive amount is added to the risk position, the risk level remains unchanged.

To illustrate how REM enables an enveloping bound on the classical monetary risk measure, and how we can construct a REM, we provide a representation of REM from the monetary risk measures.

Theorem 1. *A function $\rho:\mathcal{X}\to[0,\infty]$ is a REM if it has the representation*

$$\rho(\tilde{x})=\inf\left\{k>0\ :\ \mu_\theta(\tilde{x})\le\theta k,\ \forall\,\theta>0\right\},\tag{5}$$

where μ_θ is a class of monetary risk measures such that μ_θ is non-decreasing in $\theta>0$.

As REM is motivated by generalizing PEM, the enveloping risk measure constructed based on VaR, REM inherits the lack of convexity from VaR. However, the convexity has been recognized as a critical concern in risk management since it reflects the prevalent preference of diversification (Artzner et al. 1999, Föllmer et al. 2004). Moreover, convexity is a key property for the tractability in optimization problems. Therefore, we extend REM to incorporate the convex preference as follows.

Definition 4. A function $\rho:\mathcal{X}\to[0,\infty]$ is a Coherent Risk Enveloping Measure (CREM) if, in addition to Definition 3, it satisfies the following property for all $\tilde{x},\tilde{y}\in\mathcal{X}$:
(P5) Convexity: $\rho(\lambda\tilde{x}+(1-\lambda)\tilde{y})\le\lambda\rho(\tilde{x})+(1-\lambda)\rho(\tilde{y})$, for all $\lambda\in[0,1]$.

Similar to Theorem 1, the CREM can be dually represented by the coherent risk measure as follows.

Theorem 2. *A function $\rho:\mathcal{X}\to[0,\infty]$ is a CREM if and only if it has the representation*

$$\rho(\tilde{x})=\inf\left\{k>0\ :\ \mu_\theta(\tilde{x})\le\theta k,\ \forall\,\theta>0\right\},\tag{6}$$

where μ_θ is a class of coherent risk measure, i.e., in addition to Definition 2, it satisfies Convexity; and μ_θ is non-decreasing in $\theta>0$. Conversely, given a CREM ρ, the underlying class of coherent risk measures is given by

$$\mu_\theta(\tilde{x})=\inf_a\{a+\rho(\tilde{x}+a)\theta\}.\tag{7}$$

Similar to Theorem 1, here Theorem 2 indicates that given any class of coherent risk measure μ_θ, we can construct a corresponding CREM ρ using Equation (6). In addition, Theorem 2 implies that for any CREM ρ, there must exist a class of coherent risk measure μ_θ such that ρ can be represented by this particular class of μ_θ. In other words, CREM and coherent risk measure can be dually represented. This is a stronger result than with Theorem 1, where we cannot guarantee that all REM ρ can be represented by an underlying class of monetary risk measure.

Remark: In most situations, we can reasonably assume that the REM/CREM depends on a risk position only through its probability distribution. We call such REM/CREM *law invariant*. Formally, a REM/CREM criterion, ρ is law invariant if $\rho(\tilde{x}) = \rho(\tilde{y})$ for any $\tilde{x}$, $\tilde{y}$ with identical distribution. The property of law invariance relates closely with stochastic dominance. It is a well-known result that any law invariant coherent risk measure respects first and second stochastic dominance (see, for instance, Levy 1992). The dual representation in Theorem 2 immediately implies that so does any law invariant CREM criterion. Similarly, a law invariant REM criterion must preserve first order stochastic dominance if it can be constructed by a class of law invariant monetary risk measures in the form of Equation (1). Note that all examples of REM/CREM in this paper are law invariant. Hence, they exhibit either first order stochastic dominance (if being REM), or both first and second order stochastic dominance (if being CREM).

It is worthwhile mentioning that the CREM can also be dually represented by convex risk measures (i.e., functions satisfying Monotonicity, Translation Invariance, and Convexity and normalized by $\tilde{x} = 0$ having zero risk). Specifically, Hall et al. (2015) show that ρ is a CREM if and only if there exists a convex risk measure μ such that $\rho(\tilde{x}) = \inf\left\{\frac{1}{\alpha} \,:\, \mu(\alpha\tilde{x}) \leq 0,\ \alpha > 0\right\}$. Conversely, given a CREM ρ, the underlying normalized convex risk measure is given by $\mu(\tilde{x}) = \min\{a \,:\, \rho(\tilde{x} + a) \leq 1\}$.

Theorem 1 and Theorem 2 differ from the result in Hall et al. (2015) since we are able to provide information on the riskiness evaluated by a broad class of monetary risk measures. In particular, both Equations (5) and (6) demonstrate how REM or CREM, ρ, can serve as an enveloping bound for the monetary risk measure μ_θ. By adopting variants of some classical monetary risk measures, we can obtain corresponding enveloping bounds by Equations (5) and (6). We next provide a number of concrete examples, where the detailed reasoning is relegated to Appendix B in the online supplement.

Examples of REM and CREM

Example 1. Choose the underlying monetary risk measure as $\mu_\theta(\tilde{x}) = \mathrm{VaR}_{1-\beta(\theta)}$, where $\beta : \Re_+ \to (0,1)$ is a non-increasing function. The corresponding REM constructed by Equation (5) is indeed the PEM defined in Definition 1.

Example 2. Choose the underlying monetary risk measure as $\mu_\theta(\tilde{x}) = \mathrm{CVaR}_{1-\beta(\theta)}$, where $\beta : \Re_+ \to (0,1)$ is a non-increasing function and CVaR is defined as

$$\mathrm{CVaR}_\alpha(\tilde{x}) = \inf_\nu \left\{ \nu + \frac{1}{1-\alpha}\mathbb{E}\left[(-\nu - \tilde{x})^+\right] \right\}.$$

Because CVaR is a coherent risk measure (Rockafellar and Uryasev 2000, 2002), the corresponding REM is indeed a CREM, and it is represented as

$$\rho_{\mathrm{CVaR},\beta}(\tilde{x}) = \inf\{k > 0 \ : \ \mathrm{CVaR}_{1-\beta(\theta)}(\tilde{x}) \leq \theta k, \ \forall \theta > 0\}. \tag{8}$$

Example 3. Choose the underlying monetary risk measure as

$$\mu_\theta(\tilde{x}) = \inf_{\alpha > 0} \left\{ \alpha \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha} \right) \right] + \alpha\theta \right\},$$

which is the coherent version of the entropic risk measure $\theta \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\theta}\right)\right]$. The corresponding CREM constructed according to Equation (6) can indeed be simplified as

$$\rho_{\mathrm{Entropic}}(\tilde{x}) = \inf\left\{ k > 0 \ : \ k \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{k} \right) \right] \leq 0 \right\},$$

which coincides with the riskiness index of Aumann and Serrano (2008).

Utility-based Coherent Risk Enveloping Measure

To illustrate the relationship between CREM and the classical expected utility theory, we pay specific attention to the following CREM.

Definition 5. For any normalized convex non-decreasing function $u : \Re \to \Re$ such that $u(w) > u(0) = 1 \ \forall w > 0$, we define the utility-based CREM by

$$\rho_u(\tilde{x}) = \inf\left\{ k > 0 \ : \ \mathbb{E}\left[u\left(\frac{-\tilde{x}}{k} \right) \right] \leq 1 \right\}. \tag{9}$$

By Definition 4, we can check that the ρ_u defined by Equation (9) is a CREM. The underlying class of coherent risk measure can be constructed from Theorem 2 as

$$\begin{aligned}\mu_\theta^u(\tilde{x}) = \inf_{\alpha}\{\alpha + \rho_u(\tilde{x}+\alpha)\theta\} &= \inf_{k>0,\alpha}\left\{ \alpha + k\theta \ : \ \mathbb{E}\left[u\left(\frac{-\tilde{x}-\alpha}{k} \right) \right] \leq 1 \right\} \\ &= \inf_{t}\left\{ t \ : \ \exists k > 0 \text{ with } \mathbb{E}\left[u\left(\frac{-\tilde{x}-t}{k} + \theta \right) \right] \leq 1 \right\}.\end{aligned}$$

In addition, we have the following relationship among ρ_u, ρ_{CVaR}, and ρ_{VaR}.

Proposition 1. *Given any utility function u for which the corresponding utility-based CREM ρ_u is well defined, we always have*

$$\rho_u(\tilde{x}) \geq \rho_{\mathrm{CVaR},\beta}(\tilde{x}) \geq \rho_{\mathrm{VaR},\beta}(\tilde{x}).$$

if choosing $\beta(\theta) = 1/u(\theta) \ \forall \theta > 0$.

We can then have an enveloping bound, which can be characterized by a general group of utility functions, on the tail probability.

COROLLARY 1. *Given any $\tilde{x} \in \mathcal{X}$ with $\rho^* = \rho_u(\tilde{x}) > 0$, we have $\mathbb{P}(-\tilde{x} > \phi) \leq \frac{1}{u(\phi/\rho^*)}$, $\forall \phi > 0$.*

Based on Corollary 1, by choosing a different utility function u, we can have a different probabilistic enveloping bound.

Example 4. Let the utility function u be the classical exponential function, $u(w) = \exp(w)$. The corresponding utility-based CREM constructed by Equation (9) is actually the ρ_{Entropic} defined in Example 3. Therefore, according to Corollary 1, $\forall \tilde{x}$ with $\rho_{\text{Entropic}}(\tilde{x}) > 0$, we always have

$$\mathbb{P}\left(-\tilde{x} \geq \phi\right) \leq \frac{1}{\exp\left(\phi/\rho_{\text{Entropic}}(\tilde{x})\right)}, \ \forall \phi > 0.$$

Example 5. We choose the utility function u as the following two-piecewise linear function,

$$u_2(w) = \max\{x + 1, 0\}. \tag{10}$$

The corresponding utility-based CREM is

$$\rho_{u_2}(\tilde{x}) = \inf\left\{k > 0 \ : \ \mathbb{E}\left[u\left(-\frac{\tilde{x}}{k}\right)\right] \leq 1\right\} = \inf\left\{k > 0 \ : \ \mathbb{E}\left[\left(-\frac{\tilde{x}}{k} + 1\right)^+\right] \leq 1\right\}. \tag{11}$$

Following Corollary 1, $\forall \tilde{x} \in \mathcal{X}$, we have

$$\mathbb{P}(-\tilde{x} > \phi) \leq \frac{1}{1 + \phi/\rho_{u_2}(\tilde{x})}, \ \forall \phi > 0.$$

3. Optimizing CREM in Convex Decision Problems

In this section, we show how CREM can be incorporated with optimization, such that we can obtain the optimal decision that leads to the risk position with the lowest CREM.

To achieve explicit results for optimization and elucidate the connection with the classical expected utility theory, we pay specific attention to the utility-based CREM. Henceforth, we refer to a CREM criterion as one taking the form of Equation (9). In addition, due to the potential challenges in characterizing the true probability distribution, we use a distributionally robust optimization approach (see, for instance, Delage and Ye 2010, Wiesemann et al. 2014). Specifically, we do not assume knowledge of exact probability distribution in the probability space. Instead, we allow ambiguity in probability. We assume that the true distribution, $\mathbb{P}$ is merely known to be an element in an ambiguity set $\mathcal{P}$, which is a set of probability distributions satisfying certain constraints. In this case, the utility-based CREM ρ_u becomes

$$\rho_u(\tilde{x}) = \inf\left\{k > 0 \ : \ \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-\tilde{x}}{k}\right)\right] \leq 1\right\}. \tag{12}$$

Note that the above ρ_u actually covers the original CREM defined in Equation (9) if the ambiguity set $\mathcal{P}$ is a singleton.

Consider a general decision making problem under uncertainties. Let $v(\cdot)$ be the payoff function such that $v(\boldsymbol{y},\boldsymbol{z})$ is the payoff associated with the decision vector $\boldsymbol{y}$ when the random vector $\tilde{\boldsymbol{z}}$ is realized to be $\boldsymbol{z}$. The problem is to identify the optimal decision, $\boldsymbol{y}$ such that the CREM of $v(\boldsymbol{y},\tilde{\boldsymbol{z}})$ is minimal. The optimization problem can be formulated as

$$\begin{aligned} \rho^* = \min \ & k \\ \text{s.t.} \ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k}\right)\right] \leq 1 \\ & k > 0 \\ & \boldsymbol{y} \in \mathcal{Y}, \end{aligned} \tag{13}$$

where $\mathcal{Y}$ is a closed set and $\boldsymbol{y}\in\mathcal{Y}$ represents the deterministic restrictions on the decision $\boldsymbol{y}$. To facilitate the optimization, we make further assumption on the decision making problem.

Assumption 1. *The function $v(\boldsymbol{y},\boldsymbol{z})$ is bi-convex in $\boldsymbol{y}$ and $\boldsymbol{z}$, and can be evaluated in polynomial time. In addition, the set $\mathcal{Y}$ is convex.*

We now investigate the solution procedure of Problem (13). Note that due to the constraint of $k>0$, Problem (13) has the feasible region as an open set, which is an obstacle for optimization. Therefore, the first step is to ensure the feasible region to be a closed set. To this end, we observe the following property of the first constraint of Problem (13).

Lemma 1. *If $k^o > 0$ is such that $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k^o}\right)\right] \leq 1$, then $\forall k > k^o$ we must have $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k}\right)\right] \leq 1$.*

With the above lemma, we are ready to ensure that the feasible region of the optimization problem is a closed set. Specifically, instead of directly solving Problem (13), we analyze the decision of its ϵ-closure defined as follows.

Proposition 2. *For any $\epsilon > 0$, define ρ^*_ϵ as follows,*

$$\begin{aligned} \rho^*_\epsilon = \min \ & k \\ \text{s.t.} \ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k}\right)\right] \le 1 \\ & k \ge \epsilon \\ & \boldsymbol{y} \in \mathcal{Y}. \end{aligned} \tag{14}$$

We then have $\rho^* \le \rho^*_\epsilon \le \rho^* + \epsilon$, and $\rho^*_\epsilon = \rho^*$ if $\rho^* \ge \epsilon$.

According to Proposition 2, Problem (14) can always be made arbitrarily close to Problem (13). Practically, it is not so likely to have a decision, $\boldsymbol{y}$ with zero risk. In such cases, Problem (13) has a strictly positive optimal value, i.e., $\rho^* > 0$; hence, Problem (14) would be equivalent to Problem (13) when we choose a sufficiently small ϵ (we can keep updating by $\epsilon = \epsilon/2$ until $k = \epsilon$ is infeasible for Problem (13), in which case we have $\rho^* > \epsilon$). For the above reasons, we refer to an optimal solution of Problem (14) as one that is also optimal in Problem (13). Henceforth, we focus on Problem (14).

Based on Lemma 1, the optimal k in Problem (14) can be found by a standard bisection search on k. The search begins by initializing a search space $[\underline{k}, \overline{k}]$, where $\underline{k} = \epsilon$ and $\overline{k}$ is an arbitrary large number that is feasible for Problem (14) (we can keep doubling $\overline{k}$ until it is feasible). In each iteration of the bisection search, we consider the midpoint $k = (\underline{k} + \overline{k})/2$ and evaluate its feasibility to Problem (14) by solving the following problem,

$$\begin{aligned} \min \ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k}\right)\right] \\ \text{s.t.} \ & \boldsymbol{y} \in \mathcal{Y}. \end{aligned} \tag{15}$$

The search space is then updated by $\overline{k} = k$ if the optimal value of Problem (15) is no greater than 1, or $\underline{k} = k$ otherwise. The process repeats until the stoping criterion is satisfied.

In Problem (15), the uncertainty set, $\mathcal{P}$ is a key factor. It affects both the effectiveness in modeling uncertainties and the computational complexity. Various types of uncertainty sets have been used in the literature, such as those defined by mean and variance (Scarf 1958) and general moment information (Delage and Ye 2010). Recently, Wiesemann et al. (2014) propose using an ambiguity set that builds on expectation constraints. They demonstrate the expressiveness of the proposed ambiguity set by showing that it indeed unifies and generalizes several approaches in the literature. We follow Wiesemann et al. (2014), and let $\mathcal{P}$ be in the following standard form

$$\mathcal{P} = \left\{ \mathbb{P} = \text{proj}_{\tilde{\boldsymbol{z}}}(\mathbb{Q}) \ : \begin{array}{ll} \mathbb{E}_{\mathbb{Q}}[\boldsymbol{A}\tilde{\boldsymbol{z}} + \boldsymbol{B}\tilde{\boldsymbol{u}}] = \boldsymbol{b} & \\ \mathbb{Q}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i) \in \left[\underline{p}_i, \bar{p}_i\right] & \forall i \in \mathcal{I} \end{array} \right\}, \tag{16}$$

where $\tilde{\boldsymbol{u}}$ is an auxiliary random vector which is to empower the expressiveness, $\boldsymbol{A} \in \Re^{K\times P}$, $\boldsymbol{B} \in \Re^{K\times Q}$, $\boldsymbol{b} \in \Re^{K}$ are given parameters with P and Q being the dimension of $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{u}}$, respectively. In addition, when $\mathbb{Q}$ represents a joint probability distribution of the two random vectors $\tilde{\boldsymbol{z}}$ and $\tilde{\boldsymbol{u}}$, $\text{proj}_{\tilde{\boldsymbol{z}}}(\mathbb{Q})$ denotes the marginal distribution of $\tilde{\boldsymbol{z}}$ under $\mathbb{Q}$. Finally, we let $\mathcal{I} = \{1, 2, \ldots, I\}$; for all $i \in \mathcal{I}$, $0 \le \underline{p}_i \le \bar{p}_i \le 1$, and the sets $\mathcal{C}_i$ represent the confidence sets, which are conic representable. It has been shown that ambiguity sets $\mathcal{P}$ defined as in Equation (16) are strikingly powerful in modeling. For example, such sets can include the following information as special cases: the mean, variance, coefficient of variation, and higher-order moment information (Wiesemann et al. 2014).

To solve Problem (15) with the uncertainty set $\mathcal{P}$ defined as in Equation (16), we need the following assumptions on the confidence sets. Given any two sets A and B, we say that $A \Subset B$ if A is contained in the interior of B.

Assumption 2. 1. *For any $i, i' \in \mathcal{I}$, we have either $\mathcal{C}_i \Subset \mathcal{C}_{i'}$, $\mathcal{C}_{i'} \Subset \mathcal{C}_i$, or $\mathcal{C}_i \cap \mathcal{C}_{i'} = \emptyset$.*

2. *The confidence set $\mathcal{C}_I$ is bounded and with probability one, i.e., $\underline{p}_I = \bar{p}_I = 1$.*

3. *There exists a probability distribution $\mathbb{P} \in \mathcal{P}$ such that for all i with $\underline{p}_i < \bar{p}_i$ we have $\mathbb{P}((\tilde{\boldsymbol{z}}, \tilde{\boldsymbol{u}}) \in \mathcal{C}_i) \in (\underline{p}_i, \bar{p}_i)$.*

4. *For each $i \in \mathcal{I}$, $\boldsymbol{y} \in \mathcal{Y}$ and $\theta \in \Re$, it can be verified in polynomial time whether $\max_{(\boldsymbol{z},\boldsymbol{u})\in\mathcal{C}_i} v(\boldsymbol{y}, \boldsymbol{u}) \ge \theta$.*

With the above assumption, Wiesemann et al. (2014) illustrate the techniques to incorporate the uncertainty set $\mathcal{P}$ in optimization. We can thus solve Problem (15) as follows. Given any $i \in \mathcal{I}$, denote $\mathcal{A}(i) = \{i\} \cup \{i' \in \mathcal{I} : \mathcal{C}_i \Subset \mathcal{C}_{i'}\}$.

Proposition 3. *Problem (15) is equivalent to the following problem*

$$\begin{array}{ll} \min & \boldsymbol{b}^T\boldsymbol{\zeta} + \sum_{i\in\mathcal{I}} \left(\bar{p}_i\kappa_i - \underline{p}_i\lambda_i\right) \\ \text{s.t.} & (\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^T\boldsymbol{\zeta} + \sum_{i'\in\mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) \ge u\left(\dfrac{-v(\boldsymbol{y},\boldsymbol{z})}{k}\right) \qquad \forall(\boldsymbol{z},\boldsymbol{u}) \in \mathcal{C}_i,\ \forall i \in \mathcal{I} \\ & \boldsymbol{y} \in \mathcal{Y} \\ & \boldsymbol{\kappa}, \boldsymbol{\lambda} \ge \boldsymbol{0}, \end{array} \tag{17}$$

where $\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\lambda}$ *are also decision variables.*

Note that Problem (17) has semi-infinite constraints. However, following the analysis in the supplement of Wiesemann et al. (2014), this problem is still computationally tractable when Assumptions 1 and 2 hold. Specifically, as $\mathcal{C}_i$ is conic representable, the first set of constraints can be equivalently formulated as a set of finite number of conic constraints by writing its robust counterpart.

Optimization when the underlying utility function is piecewise linear

In the rest of the paper, we consider a special case of the above decision making problem where the underlying utility function of the CREM criterion takes a specific form. In particular, we consider the utility function, u as a piecewise linear utility function as follows,

$$u(w) = \max_{n \in \mathcal{N}} \left\{ a_n w + b_n \right\}, \tag{18}$$

where $a_n \geq 0,\ b_n,\ n \in \mathcal{N} = \{1, \ldots, N\}$ are given with $a_i \neq a_j$ for all distinct $i, j \in \mathcal{N}$. The reasons for choosing the piecewise linear utility function are twofold. First, from a practical point of view, the piecewise linear function can be used to approximate any general utility function. Second, with the piecewise linear function, the optimization procedure can be simplified. We now show that the bisection search, which is used in solving Problem (14), is no longer needed in this special case.

With the utility function u taking the form given in Equation (18), Problem (14) becomes

$$\begin{aligned} \min \quad & k \\ \text{s.t.} \quad & \sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} \left[\max_{n \in \mathcal{N}} \left\{ a_n \frac{-v(\boldsymbol{y}, \tilde{\boldsymbol{z}})}{k} + b_n \right\} \right] \leq 1 \\ & k \geq \epsilon \\ & \boldsymbol{y} \in \mathcal{Y}. \end{aligned} \tag{19}$$

We now show that Problem (19) can be reformulated as follows.

PROPOSITION 4. *Problem (19) is equivalent to the following problem,*

$$
\begin{aligned}
\min\ & k \\
\text{s.t. } & \boldsymbol{b}^T\boldsymbol{\zeta} + \sum_{i\in\mathcal{I}} \left(\bar{p}_i\kappa_i - \underline{p}_i\lambda_i\right) \le k \\
& (\boldsymbol{A}\boldsymbol{z} + \boldsymbol{B}\boldsymbol{u})^T\boldsymbol{\zeta} + \sum_{i'\in\mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) \ge -a_n v(\boldsymbol{y},\boldsymbol{z}) + b_n k \qquad \forall (\boldsymbol{z},\boldsymbol{u})\in\mathcal{C}_i,\ i\in\mathcal{I},\ n\in\mathcal{N} \\
& \boldsymbol{\kappa},\boldsymbol{\lambda}\ge\boldsymbol{0} \\
& k \ge \epsilon \\
& \boldsymbol{y}\in\mathcal{Y}.
\end{aligned}
\tag{20}
$$

Similar to the discussion following Proposition 3, Problem (20) can be solved efficiently if both Assumptions 1 and 2 hold.

4. Optimization with Linear Payoff Function and Mean-Covariance Information

Here we consider a special case of the target-driven distributionally robust optimization problem and show that it can be solved efficiently even with binary decision variables. In particular, in this section, we assume the payoff function has the following linear form,

$$v(\boldsymbol{y},\tilde{\boldsymbol{z}}) = \boldsymbol{y}^T\tilde{\boldsymbol{z}}.$$

Though being special, the linear payoff function indeed includes a broad class of applications already, such as portfolio optimization, and the surgery allocation which will be discussed with details later. Moreover, the uncertainty set is characterized by the mean and covariance information of $\tilde{\boldsymbol{z}}$. Mathematically, we let

$$\mathcal{P}_M = \left\{\mathbb{P}\ :\ \mathbb{E}\left[\tilde{\boldsymbol{z}}\right] = \boldsymbol{\mu},\ \mathbb{E}\left[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^T\right] = \boldsymbol{\Sigma}\right\}.$$

4.1. Continuous Case

Our first result is on continuous case, i.e., the decision variables can vary continuously.

Theorem 3. *If $v(\boldsymbol{y},\tilde{\boldsymbol{z}})=\boldsymbol{y}^T\tilde{\boldsymbol{z}}$ and $\mathcal{P}=\mathcal{P}_M$, Problem (19) is equivalent to the following problem,*

$$\begin{aligned}
\min\ \ & k \\
\text{s.t.}\ \ & -\alpha+\delta\le k \\
& \delta+\gamma\ge\sqrt{\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}+\beta^2+(\delta-\gamma)^2} \\
& \alpha\le -a_n^2\gamma-a_n\beta+a_n\boldsymbol{y}^T\boldsymbol{\mu}-b_nk \qquad \forall n\in\mathcal{N} \\
& \delta,\gamma\ge 0 \\
& k\ge\epsilon \\
& \boldsymbol{y}\in\mathcal{Y},
\end{aligned} \tag{21}$$

which is a SOCP when $\mathcal{Y}$ is representable using SOCP constraints.

Therefore, when only continuous decision variables are involved, the problem of optimizing CREM can be reformulated as a SOCP, which can be solved efficiently.

4.2. Discrete Case

We now consider the case that $\mathcal{Y}\subseteq\{0,1\}^n$. In this case, Problem (21) is a discrete optimization problem with SOCP constraints, and hence is no longer effectively solvable. Therefore, we need an alternative approach to solve the optimal solution for the main problem (19). Unlike the continuous case, here we cannot directly solve the optimal solution. Instead, we need to follow the procedure discussed in Section 3. Specifically, we conduct a binary search on k. For each given k, we solve the following subproblem

$$\min_{\boldsymbol{y}\in\mathcal{Y}}\left\{\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\left\{-a_n\boldsymbol{y}^T\tilde{\boldsymbol{z}}+b_nk\right\}\right]\right\}. \tag{22}$$

We update the upper bound for the optimal value in Problem (19) as this k if the optimal value of Problem (22) is no greater than k, or update the lower bound as this k otherwise. We then move to the next iteration by considering k as the mid-point of the new bound. Repeat this procedure until the stopping criterion is met.

The underlying utility function has two pieces

We first focus on a simple case where the underlying piecewise linear utility function u has only two pieces.

Theorem 4. *Consider the case that $\mathcal{P}=\mathcal{P}_M$, and the underlying utility function is $u(w)=\max\{a_1w+b_1,a_2w+b_2\}$, where $b_2\le b_1=1$. Then Problem (22) has optimal value* *no greater than k if and only if the following problem, which is a quadratic optimization problem with binary decisions, is with optimal value no greater than* 0,

$$\begin{aligned} \min \;& a_1(b_2-1)k\boldsymbol{\mu}^T\boldsymbol{y}+\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \\ \text{s.t.} \;& (a_1+a_2)\,\boldsymbol{y}^T\boldsymbol{\mu}\geq(b_2-1)k \\ & \boldsymbol{y}\in\mathcal{Y}. \end{aligned} \tag{23}$$

Here the matrix $\boldsymbol{M}=\frac{(a_2-a_1)^2}{4}\boldsymbol{\Sigma}-a_1a_2\boldsymbol{\mu}\boldsymbol{\mu}^T$.

The problem (23) is a quadratic optimization problem with binary decision variables, and can be solved by solvers like CPLEX.

Multiple Pieces

When the underlying utility function has multiple pieces, in general the related optimization problems would become much more computationally demanding. Nevertheless, in this section, we propose an enumerative algorithm, such that the problem can be solved via a sequence of quadratic optimization problems with binary decision variables.

The first step is a reformulation.

Proposition 5. *Consider the case that* $\mathcal{P}=\mathcal{P}_M$. *Then there exists* $n^*\in\mathcal{N}$ *such that Problem (22) is equivalent to the following problem,*

$$\begin{aligned} \min \;& \frac{(s_1+a_{n^*})^2}{4s_2}+b_{n^*}k+s_1\boldsymbol{\mu}^T\boldsymbol{y}+s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \\ \text{s.t.} \;& 2s_1a_{n^*}+4b_{n^*}ks_2+a_{n^*}^2\geq 2s_1a_n+4b_nks_2+a_n^2 \qquad \forall n\in\mathcal{N}\setminus\{n^*\} \\ & s_2>0,\; s_1\in\Re,\; \boldsymbol{y}\in\mathcal{Y}, \end{aligned} \tag{24}$$

where $\boldsymbol{M}=\boldsymbol{\Sigma}+\boldsymbol{\mu}\boldsymbol{\mu}^T$.

However, Problem (24) only states the existence of n^* such that both Problems (22) and (24) are equivalent. It does not specify how one can obtain such a value and solve the corresponding optimization problem. A further investigation into the optimal solutions to Problem (24) is needed. We consider three cases depending on the redundance of the first set of constraints in (24).

Case 1. Among the first set of constraints in (24), no constraint is binding. In this case, the optimal solution must be a local minimizer of the optimization without considering the first set of constraints, i.e.,

$$\min_{s_2>0,\; s_1\in\Re,\; \boldsymbol{y}\in\mathcal{Y}} \left\{\frac{(s_1+a_{n^*})^2}{4s_2}+b_{n^*}k+s_1\boldsymbol{\mu}^T\boldsymbol{y}+s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}\right\}. \tag{25}$$

To solve Problem (25), we first see the objective function as a univariate function of s_2 and we must have the optimal solution $s_2 = \frac{1}{2}|s_1 + a_{n^*}|\sqrt{1/\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}}$ if $s_1 \neq -a_{n^*}$ and s_2 approaches to 0 if $s_1 = -a_{n^*}$. In both cases, the above problem becomes

$$\min_{s_1\in\Re,\ \boldsymbol{y}\in\mathcal{Y}} \left\{|s_1 + a_{n^*}|\sqrt{\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}} + b_{n^*}k + s_1\boldsymbol{\mu}^T\boldsymbol{y}\right\}.$$

— When $s_1 + a_{n^*} \geq 0$, i.e., $s_1 \geq -a_{n^*}$, the objective function is

$$s_1\left(\sqrt{\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{y}} + \boldsymbol{\mu}^T\boldsymbol{y}\right) + a_{n^*}\sqrt{\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}} + b_{n^*}k.$$

The coefficient of s_1 is strictly positive, so the optimal s_1 is the minimal, which is $-a_{n^*}$.

— When $s_1 + a_{n^*} \leq 0$, i.e., $s_1 \leq -a_{n^*}$, the objective function is

$$s_1\left(\boldsymbol{\mu}^T\boldsymbol{y} - \sqrt{\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{y}}\right) - a_{n^*}\sqrt{\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}} + b_{n^*}k.$$

The coefficient of s_1 is strictly negative, so the optimal s_1 is the maximal, which is $-a_{n^*}$.

Therefore, in this case, we can conclude that $s_1^* = -a_{n^*}$ and hence s_2^* approaches to 0. However, with this solution, the first set of constraints in (24) become $-a_{n^*}^2 \geq a_n^2 - 2a_n a_{n^*}$, i.e., $(a_{n^*} - a_n)^2 \leq 0\ \forall n \in \mathcal{N} \setminus \{n^*\}$, which must be false since we have assumed $a_i \neq a_j$ for all distinct $i, j \in \mathcal{N}$. It implies that s_1^*, s_2^* is indeed not feasible to Problem (24). Hence, Case 1 cannot happen.

Case 2. Among the first set of constraints in (24), exactly one constraint is binding. Suppose the constraint for $n = n^o$ is binding. We can solve the following problem,

$$\begin{aligned} \min\ & \frac{(s_1 + a_{n^*})^2}{4s_2} + b_{n^*}k + s_1\boldsymbol{\mu}^T\boldsymbol{y} + s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \\ \text{s.t.}\ & 2s_1a_{n^*} + 4b_{n^*}ks_2 + a_{n^*}^2 = 2s_1a_{n^o} + 4b_{n^o}ks_2 + a_{n^o}^2 \\ & s_2 > 0,\ s_1 \in \Re,\ \boldsymbol{y} \in \mathcal{Y}. \end{aligned} \tag{26}$$

It is indeed the same type of problem with the two pieces case, with the only difference that here we do not isolate the mean. We show that the result is the same.

The first constraint of Problem (26) implies

$$2\left(a_{n^*} - a_{n^o}\right)s_1 = -4k\left(b_{n^*} - b_{n^o}\right)s_2 - \left(a_{n^*}^2 - a_{n^o}^2\right)$$

$$2s_1 = -4k\frac{b_{n^*} - b_{n^o}}{a_{n^*} - a_{n^o}}s_2 - \left(a_{n^o} + a_{n^*}\right)$$

$$s_1 = -\frac{1}{2}\left(a_{n^o} + a_{n^*} + 4k\frac{b_{n^o} - b_{n^*}}{a_{n^o} - a_{n^*}}s_2\right).$$

Substituting s_1 into the objective function, we have

$$\begin{aligned}
&\frac{1}{4s_2}\left(-\frac{1}{2}\left(a_{n^o}-a_{n^*}+4k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}s_2\right)\right)^2+b_{n^*}k-\\
&\frac{1}{2}\left(a_{n^o}+a_{n^*}+4k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}s_2\right)\boldsymbol{\mu}^T\boldsymbol{y}+s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}\\
=&\frac{1}{16s_2}\left(\left(a_{n^o}-a_{n^*}\right)^2+8k\left(b_{n^o}-b_{n^*}\right)s_2+16k^2\frac{\left(b_{n^o}-b_{n^*}\right)^2}{\left(a_{n^o}-a_{n^*}\right)^2}s_2^2\right)\\
&-\frac{1}{2}\left(a_{n^o}+a_{n^*}+4k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}s_2\right)\boldsymbol{\mu}^T\boldsymbol{y}+s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}+b_{n^*}k\\
=&s_2\left(k^2\frac{\left(b_{n^o}-b_{n^*}\right)^2}{\left(a_{n^o}-a_{n^*}\right)^2}-2k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}\boldsymbol{\mu}^T\boldsymbol{y}+\left(\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)\\
&+\frac{1}{s_2}\frac{\left(a_{n^o}-a_{n^*}\right)^2}{16}-\frac{1}{2}\left(a_{n^o}+a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}+\frac{1}{2}k\left(b_{n^o}-b_{n^*}\right)+b_{n^*}k\\
=&s_2\left(\left(k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}-\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)+\\
&\frac{1}{s_2}\frac{\left(a_{n^o}-a_{n^*}\right)^2}{16}+\frac{1}{2}\left(\left(b_{n^o}+b_{n^*}\right)k-\left(a_{n^o}+a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}\right).
\end{aligned}$$

By denoting the coefficient of s_2 as A, the coefficient of $1/s_2$ as B, and the last term in above expression as C, we should see that this is indeed the univariate minimization problem in s_2, which has been tackled in the proof of Theorem 4. Therefore, the objective function can be further reduced as below

$$\frac{1}{2}\sqrt{\left(\left(k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}-\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)\left(a_{n^o}-a_{n^*}\right)^2}+\frac{1}{2}\left(\left(b_{n^o}+b_{n^*}\right)k-\left(a_{n^o}+a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}\right).$$

With the same procedure as in Theorem 4, we can verify whether this objective function is no greater than k by solving a quadratic optimization problem as follows.

LEMMA 2. *The optimal value of Problem (26) is no greater than k if and only if $f_1(n^*,n^0)\le 4k^2(1-b_{n^*})(1-b_{n^0})$, where the function $f_1(n^*,n^0)$ is defined as*

$$f_1(n^*,n^0)=\min_{\boldsymbol{y}\in\mathcal{Y}}\left\{\boldsymbol{y}^T\left(\left(a_{n^o}-a_{n^*}\right)^2\boldsymbol{\Sigma}-4a_{n^o}a_{n^*}\boldsymbol{\mu}\boldsymbol{\mu}^T\right)\boldsymbol{y}-4k\left(a_{n^o}\left(1-b_{n^*}\right)+a_{n^*}\left(1-b_{n^o}\right)\right)\boldsymbol{\mu}^T\boldsymbol{y}\right\}. \tag{27}$$

Case 3. Among the first set of constraints in (24), no less than two constraints are binding. In this case, we have no less than two equalities and only two unknown scalars s_1,s_2. Therefore, there is a chance that we can use binding constraints to solve the exact values of s_1,s_2, which we now show is always practical. Among the binding constraints, choose arbitrary two of them and denote their index as n_1, n_2, i.e., the constraints are binding for $n = n_1, n_2$. We then have

$$2\left(a_{n^*} - a_{n_1}\right) s_1 = -4k\left(b_{n^*} - b_{n_1}\right) s_2 - \left(a_{n^*}^2 - a_{n_1}^2\right)$$

$$2\left(a_{n^*} - a_{n_2}\right) s_1 = -4k\left(b_{n^*} - b_{n_2}\right) s_2 - \left(a_{n^*}^2 - a_{n_2}^2\right),$$

or

$$\begin{aligned} 2s_1 + 4k\frac{b_{n^*} - b_{n_1}}{a_{n^*} - a_{n_1}} s_2 &= -(a_{n^*} + a_{n_1}) \\ 2s_1 + 4k\frac{b_{n^*} - b_{n_2}}{a_{n^*} - a_{n_2}} s_2 &= -(a_{n^*} + a_{n_2}). \end{aligned} \tag{28}$$

Recall that we assume $a_{n_1} \neq a_{n_2}$ when choosing the utility function, it implies the RHS of the two equations are different. Hence, the LHS of the two equations are also different, which indicates that coefficients of s_2 in the two equations must not be equal. Therefore, the two linear equations are independent, and we can obtain the unique solution for s_1, s_2. Substituting s_1, s_2 into the objective function of Problem (24), it now has the only decision variables $\boldsymbol{y}$, and is a quadratic optimization as follows,

$$f_2(n^*, s_1, s_2) = \min_{\boldsymbol{y} \in \mathcal{Y}} \frac{(s_1 + a_{n^*})^2}{4s_2} + b_{n^*}k + s_1 \boldsymbol{\mu}^T \boldsymbol{y} + s_2 \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y}. \tag{29}$$

It can then be solved by standard solvers.

Based on the above results, we propose Algorithm 1: Enumerative Algorithm, for verifying whether the optimal value of Problem (22) is no greater than k. Recall that by Proposition 5, there exists $n^* \in \mathcal{N}$ such that the corresponding Problem (24) is equivalent to Problem (22). After that, the subsequent analysis is focusing the binding status of the first set of constraints in Problem (24), which provides no information on how n^* should be chosen. Hence, in Algorithm 1, the outermost loop is considering all possible values that n^* can take. The inner loops are considering all possible scenarios that give rise to Cases 2 and 3 (Case 1 is impossible to happen).

Theorem 5. *Consider the case that $v(\boldsymbol{y}, \tilde{\boldsymbol{z}}) = \boldsymbol{y}^T \tilde{\boldsymbol{z}}$, $\mathcal{P} = \mathcal{P}_M$. Then Problem (22) has optimal value no greater than k if and only if Enumerative Algorithm, which solves at most $N(N-1)^2$ quadratic optimization problem with binary decision variables, returns* $\text{obj} = 0$.

Algorithm 1: Enumerative Algorithm

```
Initialize obj = 1;
for n* ∈ 𝒩 do
    for n^o ∈ 𝒩 \ {n*} do
        Solve for f_1(n*, n^o) defined in (27), denote the optimal solution as y*;
        Substitute y* into Problem (26) and solve for s_1 and s_2;
        Check the feasibility of constraints
         2 s_1 a_{n*} + 4 b_{n*} k s_2 + a_{n*}^2 ≥ 2 s_1 a_n + 4 b_n k s_2 + a_n^2,  ∀n ∈ 𝒩 \ {n*, n^o};
        if above constraints are feasible and f_1(n*, n^o) ≤ 4k^2 (1 − b_{n*})(1 − b_{n^0}) then
            obj = 0;
        end
    end
    for all pairs of n_1, n_2 ∈ 𝒩 \ {n*} and n_1 ≠ n_2 do
        Solve Problem (28) for s_1, s_2;
        Check the feasibility of constraints
         2 s_1 a_{n*} + 4 b_{n*} k s_2 + a_{n*}^2 ≥ 2 s_1 a_n + 4 b_n k s_2 + a_n^2,  ∀n ∈ 𝒩 \ {n*, n_1, n_2};
        if above constraints are feasible then
            Solve for f_2(n*, s_1, s_2);
            if f_2(n*, s_1, s_2) ≤ k then
                obj = 0;
            end
        end
    end
end
return obj
```

5. Application on surgery allocation

We now apply the CREM framework on a surgery allocation problem. We first describe the problem, then provide formulations for the comparison, and finally tabulate and discuss the results.

There are I operating rooms (ORs) available for assignment and J surgery cases that need to be assigned. The decision-maker has two sets of decisions to make: (i) which OR(s) to open and (ii) the venue for each surgery case to take place in. We denote the opening cost for OR i, $i = 1, \ldots, I$ as c_i^z and let z_i be the binary decision variable such that $z_i = 1$ if OR i is open and $z_i = 0$ otherwise. Regarding the second category of decisions, let y_{ij} denote the binary decision that if case j is assigned to OR i ($y_{ij} = 1$ if it is so, and $y_{ij} = 0$ if it is not). For each assignment of surgery case to a OR, there is an associated assignment cost c_{ij}^y for all $i = 1, \ldots, I$, $j = 1, \ldots, J$.

The durations of surgeries might vary owing to various factors such as medical conditions, surgeons' experience levels, and the complexities of operations. The true distribution of the surgery duration of case j when assigned to OR i, denoted as $\tilde{t}_{ij}$, is not known but is assumed to be a member of the following ambiguity set

$$\mathcal{P}_M = \left\{ \mathbb{P} \; : \; \mathbb{E}\left[\tilde{\boldsymbol{t}}\right] = \boldsymbol{\mu}, \; \mathbb{E}\left[\left(\tilde{\boldsymbol{t}} - \boldsymbol{\mu}\right)\left(\tilde{\boldsymbol{t}} - \boldsymbol{\mu}\right)^T\right] = \boldsymbol{\Sigma} \right\},$$

where we put all $\tilde{t}_{ij}$ together as the vector $\tilde{\boldsymbol{t}}$.

Moreover, there is a time limit T_i for each OR i, i.e., the sum of realized durations of cases that are assigned to OR i should be less than or equal to T_i. Accordingly, there is a trade-off between lower opening costs with more prevalent overtime and higher opening costs with fewer overtime.

A chance-constrained formulation

Zhang et al. (2018) used a chance-constrained approach to solve the assignment problem while taking the overtime into account. They defined α as the maximum violation probability of the time limit constraint of OR i that the decision-maker finds it acceptable. The objective is to minimize the total cost, which consists of the opening and assignment costs incurred. Putting it all together, they formulated the problem as below,

$$\begin{aligned}
\min \quad & \sum_{i=1}^{I} c_i^z z_i + \sum_{i=1}^{I} \sum_{j=1}^{J} c_{ij}^y y_{ij} \\
\text{s.t.} \quad & \inf_{\mathbb{P} \in \mathcal{P}_M} \mathbb{P}\left\{ \sum_{j=1}^{J} \tilde{t}_{ij} y_{ij} \le T_i \right\} \ge 1 - \alpha \qquad\qquad i = 1, \ldots, I \\
& (z_1, \ldots, z_I, y_{11}, \ldots, y_{1J}, \ldots, y_{I1}, \ldots, y_{IJ}) \in \Xi,
\end{aligned} \tag{30}$$

where

$$\Xi = \left\{ (z_1,\ldots,z_I,y_{11},\ldots,y_{1J},\ldots,y_{I1},\ldots,y_{IJ}) \ : \ \begin{array}{ll} y_{ij} \le z_i & i=1,\ldots,I,\ j=1,\ldots,J \\ \sum_{i=1}^{I} y_{ij} = 1, & j=1,\ldots,J \\ y_{ij}, z_i \in \{0,1\}, & i=1,\ldots,I,\ j=1,\ldots,J \end{array} \right\}$$

is the set of all possible operational decisions. Specifically, in the definition of Ξ, the first set of constraints are to ensure no case will be assigned to a closed OR; the second set of constraints imply that each surgery case is assigned to exactly one open OR; the last set of constraints enforce that the decisions have to be binary.

In the above problem (30), with the first constraint, the worst-case overtime probability for each OR is no greater than the prescribed threshold, i.e., α. Indeed, the ambiguity set used by Zhang et al. (2018) coincides with the one we have discussed in Theorem 4, and their approach shares similar flavor with ours, which is to mitigate the risk of having undesired outcomes. Therefore, it is natural to consider this formulation as a benchmark. We remark that apart from Zhang et al. (2018), there are other advances in distributionally robust chance constraint problems, which are Hanasusanto et al. (2017) and Xie and Ahmed (2018). Nevertheless, the former focuses on the case where there is no multiplication between uncertainties and decisions, the latter are more on approximation, and hence we do not incorporate them as benchmark approaches.

Utility-based CREM

Now we turn our attention to the corresponding formulation which uses the CREM criterion. More specifically, we focus on the utility-based CREM, which has been extensively discussed in Sections 3 and 4. The underlying utility function is chosen to be $u(x) = \max\{x+1, 0\}$ (we have tested other utility functions and observed similar results). In the rest of this section, we simply call it utility-CREM. With this assumption and the mean-covariance ambiguity set $\mathcal{P}_M$ defined above, Theorem 4 can be used to find the optimal assignment. The feasible set $\mathcal{Y}$ is defined in a way that the last constraint of Problem (30) are included appropriately. However, an additional constraint is needed when optimizing over the utility-CREM criterion, or the comparison would not be fair at all, as one can choose to open all ORs in order to minimize the overtime risk, if there is no cost budget. Accordingly, we require our approach to return a solution with the opening and assignment costs no greater than that in Problem (30). Given the inputs T_i, c_i^z, c_{ij}^y for all $i=1,\ldots,I$, $j=1,\ldots,J$, the thresholds of violation probability α, and the ambiguity set $\mathcal{P}_M$, Problem (30) is solved first. The optimal cost to that problem, denoted as C^*, is then used as the upper bound on the opening and assignment costs when optimizing over the utility-CREM criterion.

Then the surgery allocation problem based on utility-CREM is formulated as follows,

$$
\begin{aligned}
\min\ & k \\
\text{s.t.}\ & \sum_{i=1}^{I} c_i^z z_i + \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}^y y_{ij} \le C^* \\
& \sup_{\mathbb{P}\in\mathcal{P}_M} \mathbb{E}_{\mathbb{P}}\left[\max\left\{-\frac{T_i - \sum_{j=1}^{J}\tilde{t}_{ij}y_{ij}}{k} + 1, 0\right\}\right] \le 1 \qquad i=1,\ldots,I \\
& k \ge \epsilon \\
& (z_1,\ldots,z_I,y_{11},\ldots,y_{1J},\ldots,y_{I1},\ldots,y_{IJ}) \in \Xi.
\end{aligned} \tag{31}
$$

In the above problem, the first constraint makes sure that the solution returned by the utility-CREM criterion is at least equally good as the one obtained from Problem (30) in terms of the total cost. The second and third constraints are from the definition of utility-CREM and to make sure the feasible region of the problem is a closed set. The last constraint is identical to that in Problem (30).

CVaR-based CREM

Recall that a CREM criterion can be constructed based on convex risk measures. To this end, we consider a CVaR-based CREM, which has been discussed in *Example 2* and we will simply call it CVaR-CREM in the rest of this section. Consider the problem that minimizes the CVaR-CREM of the overtime. By selecting an appropriate non-increasing function $\beta:\Re_+\to[0,1)$, the problem can be formulated as follows,

$$
\begin{aligned}
\min\ & k \\
\text{s.t.}\ & \sum_{i=1}^{I} c_i^z z_i + \sum_{i=1}^{I}\sum_{j=1}^{J} c_{ij}^y y_{ij} \le C^* \\
& \text{CVaR}_{1-\beta(\theta)}\left(T_i - \sum_{j=1}^{J}\tilde{t}_{ij}y_{ij}\right) \le \theta k, \qquad \forall\theta>0, i=1,\ldots,I \\
& k \ge \epsilon \\
& (z_1,\ldots,z_I,y_{11},\ldots,y_{1J},\ldots,y_{I1},\ldots,y_{IJ}) \in \Xi.
\end{aligned} \tag{32}
$$

Notice that all but the second constraints are identical to those in Problem (31). To connect with Problem (31), we choose $\beta(\theta) = 1/u(\theta), \forall \theta > 0$, i.e.,

$$\beta(\theta) = \frac{1}{u(\theta)} = \frac{1}{\max\{\theta + 1, 0\}}\bigg|_{\theta > 0} = \frac{1}{\theta + 1}. \tag{33}$$

Next we show that, with (33), the second set of constraints in Problem (32) can be formulated as a set of quadratic constraints for any given value of k.

Proposition 6. *For given value of k and $\beta(\theta)$ defined by (33), constraints*

$$\mathrm{CVaR}_{1-\beta(\theta)}\left(T_i - \sum_{j=1}^{J} \tilde{t}_{ij} y_{ij}\right) \le \theta k, \qquad \forall \theta > 0, i = 1, \dots, I$$

can be formulated as a set of quadratic constraints.

The formulation is presented in Appendix A in the online supplement. Hence, to solve Problem (32), we can also use a bisection search method on k and at each given value of k, we solve a quadratic optimization problem.

Numerical settings and performance comparison

With the aim of comparing the three approaches mentioned above (Problems (30), (31) and (32)), we first outline the parameters and procedures in this numerical study. Following the setup in Zhang et al. (2018) closely, we first assume $I = 6$, $J = 32$. Time limits T_i are generated from a uniform distribution, $T_i \sim U(420, 540)$, and $c_i^z = T_i^2/3600 + 3 * T_i/60$ for all $i = 1, \dots, 6$. The assignment costs c_{ij}^y for all $i = 1, \dots, 6$, $j = 1, \dots 32$ are randomly sampled from the set $\{6, 18, 30, 42\}$. Using the settings of means and standard deviations stated in Table 2, a set of 10,000 data points from truncated normal distribution is generated for each possible match between surgery cases and ORs. The ambiguity sets are then constructed based on those empirical means and covariances. Problem (30) is solved first and then the optimal cost will be used as the upper bound on the total cost when solving Problems (31) and (32). Moreover, we have $\epsilon = 0.01$ when solving Problems (31) and (32).

Once the solutions from three approaches are obtained, a different set of 100,000 data points with the same moment information contained in the ambiguity set is sampled from truncated normal distribution again. This will be used for testing the performance of three approaches. If the realized total duration of surgery cases assigned to OR i is less than the time limit T_i, the overtime of OR i is 0.

Type	1	2	3	4
Mean	50	50	25	25
Standard Deviation	50	15	25	7.5
Number of Cases	8	8	8	8

Table 2 Input parameters of the random surgery durations.

To provide a fair comparison, 10 replications are done and the reported metrics are the average value of replications. We utilize six performance measures so as to provide an all-rounded view on performances of the two approaches. The first two measures are the expectation and standard deviation of the overtime of an open OR. The third one (denoted as Prob 1) is the probability that a randomly chosen open OR is having positive overtime, whereas the fourth one (denoted as Prob 2) is the probability that at least 1 opened OR is having overtime. The prior emphasizes individual ORs, and the latter considers all open ORs as a whole. Then it is followed by the conditional average overtime, which is the expected overtime normalized by the third measure. The last metric is the worst-case overtime over all the rooms and over all sample paths.

We conduct the experiment for $\alpha = 0.05, 0.1, 0.3, 0.5, 0.7, 0.9$. For the CVaR-CREM, we set a time limit of 600 seconds for solving the subproblem, i.e., the feasibility problem of the constraints in Problem (32) for given k in the bisection search procedure, in each iteration. The results are tabulated in Table 3. A smaller value indicates a better performance for all measures included in the table. Note that all data in this section are accessible on GitHub[1].

For values of α which are not small (more specifically, $\alpha \in \{0.3, 0.5, 0.7, 0.9\}$), the two CREM approaches lead to expected overtime, standard deviation, conditional expected overtime, and worst-case overtime around 20% lower than those from the chance-constrained approach. In terms of probability in having overtime (i.e., Prob 1 and Prob 2), the CREM approaches also outperform the chance-constrained approach. Nevertheless, we acknowledge that the difference in the probability is not as substantial as in other metrics. The reason for the non-substantial advantage in Prob 1 and Prob 2 is intuitive since the overtime probability is exactly the focus of the chance-constrained model.

For smaller values of α, i.e., 0.05 and 0.1, we have to acknowledge that our CREM approaches cannot outperform the chance-constrained approach. This is not surprising

[1] See `https://github.com/vincenttf-chow/crem-surgery-allocation`

α	Approach	Expected Overtime	Std	Prob 1[*]	Prob 2[†]	Conditional Expected Overtime	Worst Case	Solving Time (s)
0.05	utility-CREM	0.00	0.05	0.00	0.00	15.29	31.15	2
	CVaR-CREM	0.00	0.08	0.00	0.00	13.21	33.82	3217
	Cha. Con[‡]	0.00	0.01	0.00	0.00	4.69	6.86	628
0.1	utility-CREM	0.01	0.57	0.08	0.00	16.49	96.04	2
	CVaR-CREM	0.04	1.08	0.18	0.00	19.41	124.99	2942
	Cha. Con	0.01	0.50	0.05	0.00	17.46	91.35	418
0.3	utility-CREM	0.47	4.29	1.78	0.25	24.86	177.50	2
	CVaR-CREM	0.47	4.26	1.76	0.24	24.65	178.16	51
	Cha. Con	0.68	5.82	2.31	0.32	29.05	215.75	57
0.5	utility-CREM	3.15	12.96	8.04	2.95	34.99	257.29	2
	CVaR-CREM	3.01	12.63	7.83	2.54	34.40	257.11	858
	Cha. Con	3.41	14.55	8.66	2.83	37.85	275.17	43
0.7	utility-CREM	6.28	19.87	14.41	6.89	40.75	294.53	2
	CVaR-CREM	6.30	19.92	14.55	6.99	40.52	305.33	682
	Cha. Con	7.39	23.14	16.26	8.14	44.94	325.07	139
0.9	utility-CREM	13.47	31.42	26.46	18.47	48.97	348.68	2
	CVaR-CREM	13.50	31.51	26.41	18.40	49.12	339.00	65
	Cha. Con	14.26	33.81	26.68	18.25	52.49	380.79	2

[*] Prob 1: Probability that a randomly chosen open OR has positive overtime.

[†] Prob 2: Probability that at least one opened OR has positive overtime.

[‡] Cha. Con: Chance-constrained model

Table 3 Comparison between CREM models and Chance-constrained approach in Surgery Assignment Problem.

since the performance criterion in Table 3 are measures of risk, while the chance-constrained approach are extremely risk averse for small α. While in the case of $\alpha = 0.1$, our utility-based CREM approach performs very similarly to the chance-constrained approach, for $\alpha = 0.05$, the risk is really neglectable since we have very low level for both the chance of having overtime (almost zero) and the worst-case overtime (around 1% to 7% of the time limit of one OR). Therefore, in such case, we can conclude that both CREM approaches and chance-constrained approach lead to solutions with sufficiently low risk.

In general, as we increase the value of α, the optimal cost of Problem (30) has a tendency to decrease since the need of risk mitigation is decreasing. Consequently, the solution space of Problem (31) becomes smaller since the upper bound on the total cost gets tightened. Interestingly, Table 3 shows that even with a smaller solution space, the CREM approaches still lead to a solution with more attractive performances than those from the chance-constrained model. However, the advantage is decreased due to the lost of flexibility implied by the smaller solution space. In summary, the CREM provides adequate protection against unfavorable outcomes while keeping the tractability of underlying problems.

To solidify the validity of the conclusion above, we have conducted another computational test that uses another distribution throughout the process. The procedure is almost identical to the previous one, except now lognormal distributions, rather than truncated normal distributions, are used. Specifically, based on the same parameters in Table 2, we generate samples using lognormal distribution and then construct the ambiguity set based on the samples. In the out-of-sample test, we also generate the out-of-samples using lognormal distributions.

As lognormal distributions are more heavy-tailed, it is expected that all the out-of-sample performances of the three approaches will deteriorate. The results are given in Table 4.

α	Approach	Expected Overtime	Std	Prob 1*	Prob 2†	Conditional Expected Overtime	Worst Case	Solving Time (s)
0.1	CREM	1.18	16.65	1.24	0.15	95.93	1683.93	2
	CVaR-CREM	1.25	16.89	1.35	0.18	93.26	1599.51	2705
	Cha. Con	1.08	16.10	1.11	0.12	97.56	1663.65	297
0.3	CREM	3.16	25.27	3.71	0.84	86.07	1698.94	2
	CVaR-CREM	3.16	25.27	3.72	0.83	85.96	1713.55	1300
	Cha. Con	3.36	26.19	3.91	0.86	86.36	1937.30	31
0.5	CREM	5.97	33.24	7.14	2.03	83.78	1750.74	2
	CVaR-CREM	5.96	33.21	7.13	2.02	83.73	1772.91	1024
	Cha. Con	6.37	34.57	7.56	2.21	84.02	1888.82	37
0.7	CREM	10.29	42.67	12.52	4.90	82.41	1864.16	2
	CVaR-CREM	10.32	42.72	12.58	4.95	82.10	1861.41	681
	Cha. Con	10.99	44.87	13.08	5.04	84.11	2019.24	81
0.9	CREM	14.75	50.27	18.20	9.12	81.38	2016.04	2
	CVaR-CREM	14.73	50.32	18.10	9.01	81.65	2018.32	57
	Cha. Con	16.24	53.64	19.29	9.43	85.15	2098.60	1

* Prob 1: Probability that a randomly chosen open OR has positive overtime.

† Prob 2: Probability that at least one opened OR has positive overtime.

‡ Cha. Con: Chance-constrained model

Table 4 Comparison between CREM model and Chance-constrained approach in Surgery Assignment Problem with Lognormal Distributions.

Notice that the chance-constrained approach could not return an optimal solution within 2 hours for $\alpha = 0.05$. Therefore, unlike Table 3, here in Table 4 we do not include the result for the case of $\alpha = 0.05$. By the results in Table 4, the two CREM approaches in general dominate the chance-constraint approach under lognormal distributions. The performances of the three approaches are close when α is relatively small, i.e., $\alpha = 0.1$, whereas the edges of the CREM are more obvious as we increase the value of α.

6. Concluding Remarks

In this paper, we propose the Risk Enveloping Measure (REM) to help decision makers evaluate the risk from a broad perspective. For example, the Probabilistic Enveloping Measure (PEM), which is a special case of REM, carries information for the bound on tails probability at all levels. Generalizing from PEM, REM criterion can build an enveloping bound for any monetary risk measure at all levels of risk aversion. We then study the coherent version of REM (CREM), and develop the dual representation between CREM and the classical coherent risk measure.

We incorporate the CREM framework with distributionally robust optimization, and provide an efficient method for solving the optimal solution. Applying the CREM criterion to the surgery assignment problem, we obtain a formulation which can be solved efficiently. Comparing the formulation to the selected literature which also mitigates the overtime risk in a distributional robust setting, our numerical study suggests that our CREM model provides an interesting alternative for regulating risk.

Appendix A: Proof of statements

Theorem 1 *A function $\rho:\mathcal{X}\to[0,\infty]$ is a REM if it has the representation*

$$\rho(\tilde{x})=\inf\{k>0\ :\ \mu_\theta(\tilde{x})\le\theta k,\ \forall\,\theta>0\},$$

where μ_θ is a class of monetary risk measures such that μ_θ is non-decreasing in $\theta>0$.

Proof of Theorem 1

We complete the proof by showing that the ρ defined by Equation (5) satisfies all of the properties in Definition 3. The property of Monotonicity is trivial. To show Positive Homogeneity, observe that $\forall k>0$,

$$\begin{aligned}\rho(k\tilde{x}) &= \inf\{\alpha>0\ :\ \mu_\theta(k\tilde{x})\le\theta\alpha,\ \forall\,\theta>0\}\\ &= \inf\{k\beta>0\ :\ k\mu_\theta(\tilde{x})\le\theta k\beta,\ \forall\,\theta>0\}\\ &= k\inf\{\beta>0\ :\ \mu_\theta(\tilde{x})\le\theta\beta,\ \forall\,\theta>0\}\\ &= k\rho(\tilde{x}),\end{aligned}$$

where the second equality follows from the replacement of α with $k\beta$ and Positive Homogeneity of μ_θ.

We now show Satisficing. For the case that $\tilde{x}\ge 0$, $\forall k,\theta>0$, we have $\mu_\theta(\tilde{x})\le\mu_\theta(0)=0\le k\theta$ and hence $\rho(\tilde{x})=0$. When $\tilde{x}<0$, we can have $\epsilon<0$ such that $\mathbb{P}(\tilde{x}\le\epsilon)=1$; $\forall k>0$, choose $\theta=-\epsilon/2k>0$ and hence $\mu_\theta(\tilde{x})\ge\mu_\theta(\epsilon)=-\epsilon>\theta k$, and $\rho(\tilde{x})=\infty$.

We prove the Right Continuity in three scenarios. Note that $\rho(\tilde{x}+a)$ is non-decreasing when a is decreasing, and $\rho(\tilde{x}+a)\le\rho(\tilde{x})$ for $a>0$. Hence, $\lim_{a\downarrow 0}\rho(\tilde{x}+a)$ always exists.

- Consider the case of $\rho(\tilde{x})=0$. As $0\le\rho(\tilde{x}+a)\le\rho(\tilde{x})=0$ $\forall a>0$, $\lim_{a\downarrow 0}\rho(\tilde{x}+a)=0$.
- Consider the case of $\rho(\tilde{x})\in(0,\infty)$. To prove $\lim_{a\downarrow 0}\rho(\tilde{x}+a)=\rho(\tilde{x})$, we consider any $\epsilon\in(0,\rho(\tilde{x}))$, and we need to show that $\exists\bar{a}>0$ such that $\forall a\in(0,\bar{a})$, $\rho(\tilde{x}+a)\ge\rho(\tilde{x})-\epsilon$. Note that by the representation of Equation (5), $\rho(\tilde{x})=\inf\{k<0:\mu_\theta(\tilde{x})\le\theta k,\forall\theta>0\}\in(0,\infty)$. Hence, we must have $\theta^*>0$ such that $\mu_{\theta^*}(\tilde{x})>\theta^*(\rho(\tilde{x})-\epsilon)$. Choose $\bar{a}=\mu_{\theta^*}(\tilde{x})-\theta^*(\rho(\tilde{x})-\epsilon)$. For any $a\in(0,\bar{a})$,

$$\mu_{\theta^*}(\tilde{x}+a)=\mu_{\theta^*}(\tilde{x})-a>\mu_{\theta^*}(\tilde{x})-\bar{a}=\theta^*(\rho(\tilde{x})-\epsilon).$$

Therefore, we have $\rho(\tilde{x}+a)\ge\rho(\tilde{x})-\epsilon$ according to the representation (5).

- Consider the case of $\rho(\tilde{x})=\infty$. Assume to the contrary that $k^*=\lim_{a\downarrow 0}\rho(\tilde{x}+a)<\infty$. Consider any small positive ϵ. By the definition of the limit, we have $\bar{a}>0$ such that $\forall a\in(0,\bar{a})$, $\rho(\tilde{x}+a)\le k^*+\epsilon$. In contrast, observe that $\rho(\tilde{x})=\inf\{k>0:\mu_\theta(\tilde{x})\le\theta k,\forall\theta>0\}=\infty$; then, we must have $\theta^*>0$ such that $\mu_{\theta^*}(\tilde{x})>\theta^*(k^*+2\epsilon)$. Let $\Delta=\mu_{\theta^*}(\tilde{x})-\theta^*(k^*+2\epsilon)>0$, and choose $a=\min\{\frac{\bar{a}}{2},\frac{\Delta}{2}\}$. We then have $\rho(\tilde{x}+a)\le k^*+\epsilon$ because $a\in(0,\bar{a})$. However,

$$\mu_{\theta^*}(\tilde{x}+a)=\mu_{\theta^*}(\tilde{x})-a\ge\mu_{\theta^*}(\tilde{x})-\frac{\Delta}{2}=\mu_{\theta^*}(\tilde{x})-\Delta+\frac{\Delta}{2}=\theta^*(k^*+2\epsilon)+\frac{\Delta}{2}>\theta^*(k^*+2\epsilon),$$

contradicts with $\inf\{k>0:\mu_\theta(\tilde{x}+a)\le\theta k,\forall\theta>0\}=\rho(\tilde{x}+a)\le k^*+\epsilon$. Therefore, the assumption is false, $\lim_{a\downarrow 0}\rho(\tilde{x}+a)=\infty$.

Q.E.D.

Theorem 2 *A function $\rho:\mathcal{X}\to[0,\infty]$ is a CREM if and only if it has the representation*

$$\rho(\tilde{x})=\inf\{k>0\ :\ \mu_\theta(\tilde{x})\le\theta k,\ \forall\,\theta>0\},$$

where μ_θ is a class of coherent risk measure, i.e., in addition to Definition 2, it satisfies Convexity; and μ_θ is non-decreasing in $\theta>0$. Conversely, given a CREM ρ, the underlying class of coherent risk measures is given by

$$\mu_\theta(\tilde{x})=\inf_a\{a+\rho(\tilde{x}+a)\theta\}.$$

Proof of Theorem 2

"$\Rightarrow$" We first prove the "if" direction. As we have Theorem 1 already, here it suffices to show that the ρ defined by Equation (6) has Convexity when μ_θ is a coherent risk measure. To this end, consider any $\lambda\in(0,1)$, $\tilde{x},\tilde{y}\in\mathcal{X}$ with $\rho(\tilde{x}),\rho(\tilde{y})<\infty$. Observe that given any $k>\lambda\rho(\tilde{x})+(1-\lambda)\rho(\tilde{y})$, there exists $m_x>\rho(\tilde{x})$, $m_y>\rho(\tilde{y})$ such that $k=\lambda m_x+(1-\lambda)m_y$. Hence, $\forall\theta>0$,

$$\begin{aligned}\mu_\theta(\lambda\tilde{x}+(1-\lambda)\tilde{y}) &\le \lambda\mu_\theta(\tilde{x})+(1-\lambda)\mu_\theta(\tilde{y})\\ &\le \lambda\theta m_x+(1-\lambda)\theta m_y\\ &= \theta k,\end{aligned}$$

where the first equality is due to Convexity of μ_θ, the second equality follows from the representation (6) and $m_x>\rho(\tilde{x})$, $m_y>\rho(\tilde{y})$. Therefore, $\rho(\lambda\tilde{x}+(1-\lambda)\tilde{y})\le\lambda\rho(\tilde{x})+(1-\lambda)\tilde{y}$.

"$\Leftarrow$" We now prove the "only if" direction. First, we show that the μ_θ defined by Equation (7) is a class of coherent risk measures that is non-decreasing in θ. The non-decreasing property in θ is trivial. The property of Monotonicity of μ_θ is also straightforward. To show Positive Homogeneity, observe that $\forall k>0$,

$$\begin{aligned}\mu_\theta(k\tilde{x}) &= \inf_a\{a+\rho(k\tilde{x}+a)\theta\}\\ &= \inf_a\{a+k\rho(\tilde{x}+a/k)\theta\}\\ &= \inf_\beta\{k\beta+k\rho(\tilde{x}+\beta)\theta\}\\ &= k\inf_\beta\{\beta+\rho(\tilde{x}+\beta)\theta\}\\ &= k\mu_\theta(\tilde{x}),\end{aligned}$$

where the second equality follows from the Positive Homogeneity of ρ, and the third equality just follows from the replacement of a with $k\beta$. To prove Translation Invariance, notice that $\forall c\in\Re$,

$$\begin{aligned}\mu_\theta(\tilde{x}+c) &= \inf_a\{a+\rho(\tilde{x}+c+a)\theta\}\\ &= \inf_\beta\{\beta-c+\rho(\tilde{x}+\beta)\theta\}\\ &= \inf_\beta\{\beta+\rho(\tilde{x}+\beta)\theta\}-c\\ &= \mu_\theta(\tilde{x})-c,\end{aligned}$$

where the second equality just follows from the replacement of a with $\beta - c$. To show Convexity, consider any $a_x, a_y \in \Re$,

$$\begin{aligned}&\lambda(a_x + \rho(\tilde{x} + a_x)\theta) + (1-\lambda)(a_y + \rho(\tilde{y} + a_y)\theta)\\ \geq\ &(\lambda a_x + (1-\lambda)a_y) + \rho(\lambda\tilde{x} + (1-\lambda)\tilde{y} + \lambda a_x + (1-\lambda)a_y)\theta\\ \geq\ &\mu_\theta(\lambda\tilde{x} + (1-\lambda)\tilde{y}),\end{aligned}$$

where the two inequalities follow from Convexity of ρ, and the representation (7), respectively. Taking the infimum over a_x, a_y, we have Convexity of μ_θ.

To close the loop, we next show that the right-hand-side (RHS) of Equation (6) is also ρ when μ_θ is given in the form of (7). Denote k^* as RHS of Equation (6). When μ_θ is given in the form of (7),

$$\begin{aligned}k^* &= \inf\{k > 0\ :\ \mu_\theta(\tilde{x}) \leq \theta k,\ \forall\theta > 0\}\\ &= \inf\{k > 0\ :\ \forall\theta > 0,\ \inf_a\{a + \rho(\tilde{x} + a)\theta\} \leq \theta k\}\\ &= \inf\{k > 0\ :\ \forall\theta > 0,\ \inf_a\{a/\theta + \rho(\tilde{x} + a)\} \leq k\}\\ &= \max\{0,\ k^o\},\end{aligned}$$

where the third equality holds for Positive Homogeneity of ρ, and we denote

$$k^o = \sup_{\theta>0}\inf_a\{a/\theta + \rho(\tilde{x} + a)\}. \tag{34}$$

We first notice that

$$k^o = \sup_{\theta>0}\inf_a\{a/\theta + \rho(\tilde{x} + a)\} \leq \sup_{\theta>0}\{0/\theta + \rho(\tilde{x} + 0)\} = \rho(\tilde{x}). \tag{35}$$

We now complete the proof by showing $k^* = \rho(\tilde{x})$ in all three possible scenarios.

1. Consider the case that $\rho(\tilde{x}) = 0$. According to Equation (35), $k^o \leq 0$, $k^* = \max\{0, k^o\} = 0 = \rho(\tilde{x})$.
2. Consider the case that $\rho(\tilde{x}) = \infty$. In this case, when $\theta = 1$,

$$\begin{aligned}\inf_a\{a/\theta + \rho(\tilde{x} + a)\} &= \inf_a\{a + \rho(\tilde{x} + a)\}\\ &= \inf_{a>0}\{a + \rho(\tilde{x} + a)\}\\ &> 0,\end{aligned}$$

where the second equality is due to that $\rho(\tilde{x} + a) = \infty\ \forall a \leq 0$, and the inequality holds because Right Continuity of ρ implies $\lim_{a\downarrow 0}\rho(\tilde{x} + a) = \rho(\tilde{x}) = \infty$. Hence, we have

$$k^o = \sup_{\theta>0}\inf_a\{a/\theta + \rho(\tilde{x} + a)\} \geq \inf_a\{a/1 + \rho(\tilde{x} + a)\} > 0.$$

We now prove $k^o = \infty$ by contradiction. Assume to the contrary, i.e., $k^o \in (0, \infty)$. Due to $\rho(\tilde{x}) = \infty$ and Right Continuity, $\exists\bar{a} \in (0, \infty)$ such that $\forall a \in (0, \bar{a})$, $\rho(\tilde{x} + a) = \infty$. Choose $\theta^* = \bar{a}/2k^o \in (0, \infty)$. Note that $\forall a \leq \bar{a}$, $a/\theta^* + \rho(\tilde{x} + a) = \infty$ because $\rho(\tilde{x} + a) = \infty$; $\forall a > \bar{a}$, $a/\theta^* + \rho(\tilde{x} + a) \geq a/\theta^* > \bar{a}/\theta^* = 2k^o$. Hence,

$$k^o = \sup_{\theta>0}\inf_a\{a/\theta + \rho(\tilde{x} + a)\} \geq \inf_a\{a/\theta^* + \rho(\tilde{x} + a)\} = \inf_{a>\bar{a}}\{a/\theta^* + \rho(\tilde{x} + a)\} \geq 2k^o,$$

which contradicts with the equation $k^o \in (0, \infty)$. Therefore, the assumption is false, $k^o = \infty$, and hence $k^* = \infty$.

3. Consider the case that $\rho(\tilde{x}) \in (0,\infty)$. Due to the Monotonicity and Convexity of ρ, we have that $f(a) = \rho(\tilde{x}+a)$ is convex and non-increasing in a. Hence, we can find $d < 0$ such that $d \in \partial f(0)$ where ∂ denotes the subdifferential. Choose $\theta^* = -1/d > 0$. Then $0 = 1/\theta^* + d \in \partial g(0)$ where we denote $g(a) = a/\theta^* + f(a) = a/\theta^* + \rho(\tilde{x}+a)$. Thus, 0 minimizes the convex function $g(a)$,

$$k^o = \sup_{\theta>0} \inf_a \{a/\theta + \rho(\tilde{x}+a)\} \geq \inf_a \{a/\theta^* + \rho(\tilde{x}+a)\} = \inf_a g(a) = g(0) = \rho(\tilde{x}),$$

where the last equality follows from the fact that 0 minimizes $g(a)$. Together with Equation (35), we have $k^o = \rho(\tilde{x})$. Hence, $k^* = \rho(\tilde{x})$.

In conclusion, we always have $k^* = \rho(\tilde{x})$. Q.E.D.

Proposition 1 *Given any utility function u for which the corresponding utility-based CREM ρ_u is well defined, we always have*

$$\rho_u(\tilde{x}) \geq \rho_{\mathrm{CVaR},\beta}(\tilde{x}) \geq \rho_{\mathrm{VaR},\beta}(\tilde{x}).$$

if choosing $\beta(\theta) = 1/u(\theta)$ $\forall \theta > 0$.

Proof of Proposition 1

The second inequality is straightforward because $\mathrm{CVaR}_\alpha(\tilde{x}) \geq \mathrm{VaR}_\alpha(\tilde{x})$ for all $\alpha \in (0,1)$. We now just focus on the first inequality. To this end, it suffices to show that the set in the RHS of Equation (9) is a subset of that on the RHS of Equation (8). Consider any $k > 0$ such that $\mathbb{E}[u(-\tilde{x}/k)] \leq 1$, and any $\theta > 0$. As u is non-decreasing and convex and $u(w) > u(0)$ $\forall w > 0$, we can find $b \in \partial u(\theta) > 0$. Denote $a = u(\theta)/b > 0$. We then have $\forall x \in \Re$, $u(x) \geq u(\theta) + b(x-\theta)$, i.e., $\max\{0, 1 + (x-\theta)/a\} \leq u(x)/u(\theta)$. This implies

$$\mathbb{E}\left[\left(1 + \frac{\frac{-\tilde{x}}{k} - \theta}{a}\right)^+\right] = \mathbb{E}\left[\max\left\{0, 1 + \frac{\frac{-\tilde{x}}{k} - \theta}{a}\right\}\right] \leq \frac{\mathbb{E}\left[u\left(\frac{-\tilde{x}}{k}\right)\right]}{u(\theta)} \leq \frac{1}{u(\theta)}.$$

Therefore,

$$\frac{1}{\beta(\theta)}\mathbb{E}\left[(-\tilde{x} - k\theta + ak)^+\right] = \frac{1}{1/u(\theta)}\mathbb{E}\left[(-\tilde{x} - k\theta + ak)^+\right] \leq ak,$$

$$\inf_{v\in\Re}\left\{v + \frac{1}{\beta(\theta)}\mathbb{E}\left[(-\tilde{x} - v)^+\right]\right\} \leq (k\theta - ak) + \frac{1}{\beta(\theta)}\mathbb{E}\left[(-\tilde{x} - (k\theta - ak))^+\right] \leq (k\theta - ak) + ak = k\theta.$$

In other words, $\mathrm{CVaR}_{1-\beta(\theta)}(\tilde{x}) \leq k\theta$ $\forall \theta > 0$, k belongs to the set in the RHS of Equation (8). Q.E.D.

Corollary 1 *Given any $\tilde{x} \in \mathcal{X}$ with $\rho^* = \rho_u(\tilde{x}) > 0$, we have $\mathbb{P}(-\tilde{x} > \phi) \leq \frac{1}{u(\phi/\rho^*)}$, $\forall \phi > 0$.*

Proof of Corollary 1

It is straightforward from Equation (3), Proposition 1 and the fact that $\rho_{\mathrm{VaR},\beta}(\tilde{x}) = \mathrm{PEM}_\beta(\tilde{x})$. Q.E.D.

Lemma 1 *If $k^o > 0$ is such that $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k^o}\right)\right] \leq 1$, then $\forall k > k^o$ we must have $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k}\right)\right] \leq 1$.*

Proof of Lemma 1

To simplify the notation, we denote $\tilde{x} = v(\boldsymbol{y}, \tilde{\boldsymbol{z}})$. We prove the lemma by contradiction. Assume the contrary, i.e., $\exists k^u > k^o$ such that $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-\tilde{x}}{k^u}\right)\right] > 1$. We then have

$$\begin{aligned}
\frac{k^o}{k^u}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-\tilde{x}}{k^o}\right)\right]+\left(1-\frac{k^o}{k^u}\right) &= \frac{k^o}{k^u}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-\tilde{x}}{k^o}\right)\right]+\left(1-\frac{k^o}{k^u}\right)u(0)\\
&= \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[\frac{k^o}{k^u}u\left(\frac{-\tilde{x}}{k^o}\right)+\left(1-\frac{k^o}{k^u}\right)u(0)\right]\\
&\geq \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[u\left(\frac{k^o}{k^u}\cdot\frac{-\tilde{x}}{k^o}+\left(1-\frac{k^o}{k^u}\right)\cdot 0\right)\right]\\
&= \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-\tilde{x}}{k^u}\right)\right]\\
&> 1,
\end{aligned}$$

where the first equality follows from the normalization of $u(0) = 1$, and the inequality is due to the convexity of u. It implies $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-\tilde{x}}{k^o}\right)\right] > 1$, and contradicts with the original condition of $\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k^o}\right)\right] \leq 1$. Therefore, the assumption is wrong and the proof is completed. Q.E.D.

Proposition 2 *For any $\epsilon > 0$, define ρ^*_ϵ as follows,*

$$\begin{aligned}
\rho^*_\epsilon = \min\ & k\\
\text{s.t.}\ & \sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[u\left(\frac{-v(\boldsymbol{y},\tilde{\boldsymbol{z}})}{k}\right)\right]\leq 1\\
& k\geq\epsilon\\
& \boldsymbol{y}\in\mathcal{Y}.
\end{aligned}$$

We then have $\rho^* \leq \rho^*_\epsilon \leq \rho^* + \epsilon$, and $\rho^*_\epsilon = \rho^*$ if $\rho^* \geq \epsilon$.

Proof of Proposition 2

Note that $\rho^* \leq \rho^*_\epsilon$ is straightforward because the feasible set of Problem (14) is a subset of that from Problem (13). According to Lemma 1, with every feasible solution $(k, \boldsymbol{y})$ in Problem (13), we can construct a solution $(k + \epsilon, \boldsymbol{y})$ that is feasible in Problem (14) and with an objective value only increased by ϵ. Therefore, we have $\rho^*_\epsilon \leq \rho^* + \epsilon$. Finally, if $\rho^* \geq \epsilon$, the two feasible sets of Problem (13) and Problem (14) are identical; and hence, we have $\rho^*_\epsilon = \rho^*$. Q.E.D.

Proposition 3 *Problem (15) is equivalent to the following problem*

$$\begin{aligned}
\min\ & \boldsymbol{b}^T\boldsymbol{\zeta}+\sum_{i\in\mathcal{I}}\left(\bar{p}_i\kappa_i-\underline{p}_i\lambda_i\right)\\
\text{s.t.}\ & (\boldsymbol{A}\boldsymbol{z}+\boldsymbol{B}\boldsymbol{u})^T\boldsymbol{\zeta}+\sum_{i'\in\mathcal{A}(i)}(\kappa_{i'}-\lambda_{i'})\geq u\left(\frac{-v(\boldsymbol{y},\boldsymbol{z})}{k}\right) \qquad \forall(\boldsymbol{z},\boldsymbol{u})\in\mathcal{C}_i,\ \forall i\in\mathcal{I}\\
& \boldsymbol{y}\in\mathcal{Y}\\
& \boldsymbol{\kappa},\boldsymbol{\lambda}\geq\boldsymbol{0},
\end{aligned}$$

where $\boldsymbol{y}, \boldsymbol{\zeta}, \boldsymbol{\kappa}, \boldsymbol{\lambda}$ are also decision variables.

Proof of Proposition 3

It follows immediately from Theorem 1 in the supplement of Wiesemann et al. (2014). Q.E.D.

Proposition 4 *Problem (19) is equivalent to the following problem,*

$$\begin{array}{ll} \min & k \\ \text{s.t.} & \boldsymbol{b}^T\boldsymbol{\zeta} + \sum_{i\in\mathcal{I}} \left(\bar{p}_i\kappa_i - \underline{p}_i\lambda_i\right) \le k \\ & (\boldsymbol{A}\boldsymbol{z}+\boldsymbol{B}\boldsymbol{u})^T\boldsymbol{\zeta} + \sum_{i'\in\mathcal{A}(i)} (\kappa_{i'} - \lambda_{i'}) \ge -a_n v(\boldsymbol{y},\boldsymbol{z}) + b_n k \qquad \forall(\boldsymbol{z},\boldsymbol{u})\in\mathcal{C}_i,\ i\in\mathcal{I},\ n\in\mathcal{N} \\ & \boldsymbol{\kappa},\boldsymbol{\lambda}\ge \boldsymbol{0} \\ & k\ge\epsilon \\ & \boldsymbol{y}\in\mathcal{Y}. \end{array}$$

Proof of Proposition 4

To complete the proof, it suffices to show the equivalence between the first constraint in Problem (19) and the first two constraints in Problem (20). To this end, observe that the former can be reformulated as

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\{-a_n v(\boldsymbol{y},\tilde{\boldsymbol{z}}) + b_n k\}\right] \le k$$

by multiplying both sides by k. Similar to Proposition 3, the left-hand-side of the above inequality is indeed the optimal value of

$$\begin{array}{ll} \min & \boldsymbol{b}^T\boldsymbol{\zeta} + \sum_{i\in\mathcal{I}}\left(\bar{p}_i\kappa_i - \underline{p}_i\lambda_i\right) \\ \text{s.t.} & (\boldsymbol{A}\boldsymbol{z}+\boldsymbol{B}\boldsymbol{u})^T\boldsymbol{\zeta} + \sum_{i'\in\mathcal{A}(i)}(\kappa_{i'}-\lambda_{i'}) \ge -a_n v(\boldsymbol{y},\boldsymbol{z}) + b_n k \qquad \forall(\boldsymbol{z},\boldsymbol{u})\in\mathcal{C}_i,\ i\in\mathcal{I},\ n=1,\dots,N. \end{array}$$

Q.E.D.

Theorem 3 *If* $v(\boldsymbol{y},\tilde{\boldsymbol{z}}) = \boldsymbol{y}^T\tilde{\boldsymbol{z}}$ *and* $\mathcal{P}=\mathcal{P}_M$, *Problem (19) is equivalent to the following problem,*

$$\begin{array}{ll} \min & k \\ \text{s.t.} & \delta+\gamma \ge \sqrt{\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} + \beta^2 + (\delta-\gamma)^2} \\ & \alpha \le -a_n^2\gamma - a_n\beta + a_n\boldsymbol{y}^T\boldsymbol{\mu} - b_n k \qquad \forall n\in\mathcal{N} \\ & \delta,\gamma\ge 0 \\ & k\ge\epsilon \\ & \boldsymbol{y}\in\mathcal{Y}, \end{array}$$

which is a SOCP when $\mathcal{Y}$ *is representable using SOCP constraints.*

Proof of Theorem 3

Denote $\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{z}} - \boldsymbol{\mu}$, i.e., $\tilde{\boldsymbol{z}} = \tilde{\boldsymbol{w}} + \boldsymbol{\mu}$. For the first constraint in Problem (19), multiplying k on both sides, then the LHS is equivalent to

$$\sup_{\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{w}}]=\boldsymbol{0},\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{w}}\tilde{\boldsymbol{w}}^T]=\boldsymbol{\Sigma}} \mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\left\{-a_n\boldsymbol{y}^T\tilde{\boldsymbol{w}} - a_n\boldsymbol{y}^T\boldsymbol{\mu} + b_n k\right\}\right]. \tag{36}$$

According to the projection theorem by Popescu (2007), the value of equation (36) equals to

$$\sup_{\mathbb{E}_{\mathbb{P}}[\tilde{x}]=0,\ \mathbb{E}_{\mathbb{P}}[\tilde{x}^2]=\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}} \mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\left\{-a_n\tilde{x} - a_n\boldsymbol{y}^T\boldsymbol{\mu} + b_n k\right\}\right],$$

which is

$$\begin{array}{ll} \max & \mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\left\{-a_n\tilde{x} - a_n\boldsymbol{y}^T\boldsymbol{\mu} + b_n k\right\}\right] \\ \text{s.t.} & \mathbb{E}_{\mathbb{P}}[\tilde{x}] = 0 \\ & \mathbb{E}_{\mathbb{P}}[\tilde{x}^2] = \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} \\ & \mathbb{E}_{\mathbb{P}}[1] = 1. \end{array}$$

Its duality is

$$\begin{array}{ll} \min & s_0 + s_2 \boldsymbol{y}^T \boldsymbol{\Sigma} \boldsymbol{y} \\ \text{s.t.} & s_0 + s_1 x + s_2 x^2 \geq -a_n x - a_n \boldsymbol{y}^T \boldsymbol{\mu} + b_n k \qquad \forall x \in \Re,\ n \in \mathcal{N}. \end{array} \tag{37}$$

The constraint in the problem (37) actually implies $s_2 > 0$, and can be reformulated as

$$s_0 + a_n \boldsymbol{y}^T \boldsymbol{\mu} - b_n k \geq \max_x \left\{ -s_2 x^2 + (-a_n - s_1) x \right\} = \frac{(s_1 + a_n)^2}{4 s_2} \qquad \forall n \in \mathcal{N}. \tag{38}$$

Therefore, the problem (37) is

$$\begin{array}{ll} \min & s_0 + s_2 \boldsymbol{y}^T \boldsymbol{\Sigma} \boldsymbol{y} \\ \text{s.t.} & \dfrac{(s_1 + a_n)^2}{4 s_2} \leq s_0 + a_n \boldsymbol{y}^T \boldsymbol{\mu} - b_n k \qquad \forall n \in \mathcal{N} \\ & s_2 > 0. \end{array} \tag{39}$$

We change the decision variables as follows,

$$\begin{aligned} s_0 &= -\alpha + \frac{\beta^2}{4\gamma} \\ s_1 &= \frac{\beta}{2\gamma} \\ s_2 &= \frac{1}{4\gamma}. \end{aligned}$$

The problem (39) is then changed to

$$\begin{array}{ll} \min & -\alpha + \dfrac{\beta^2}{4\gamma} + \dfrac{\boldsymbol{y}^T \boldsymbol{\Sigma} \boldsymbol{y}}{4\gamma} \\ \text{s.t.} & \dfrac{(\beta/2\gamma + a_n)^2}{1/\gamma} \leq -\alpha + \dfrac{\beta^2}{4\gamma} + a_n \boldsymbol{y}^T \boldsymbol{\mu} - b_n k \qquad \forall n \in \mathcal{N} \\ & \gamma > 0, \end{array} \tag{40}$$

which is equivalent to

$$\begin{array}{ll} \min & -\alpha + \delta \\ \text{s.t.} & \delta \geq \dfrac{\beta^2 + \boldsymbol{y}^T \boldsymbol{\Sigma} \boldsymbol{y}}{4\gamma} \\ & \alpha \leq -a_n^2 \gamma - a_n \beta + a_n \boldsymbol{y}^T \boldsymbol{\mu} - b_n k \qquad \forall n \in \mathcal{N} \\ & \gamma > 0. \end{array} . \tag{41}$$

The original problem then becomes Problem (21). Q.E.D.

Theorem 4 *Consider the case that $\mathcal{P} = \mathcal{P}_M$, and the underlying utility function is $u(w) = \max\{a_1 w + b_1, a_2 w + b_2\}$, where $b_2 \leq b_1 = 1$. Then Problem (22) has optimal value no greater than k if and only if the following problem, which is a quadratic optimization problem with binary decisions, is with optimal value no greater than 0,*

$$\begin{array}{ll} \min & a_1 (b_2 - 1) k \boldsymbol{\mu}^T \boldsymbol{y} + \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y} \\ \text{s.t.} & (a_1 + a_2)\, \boldsymbol{y}^T \boldsymbol{\mu} \geq (b_2 - 1) k \\ & \boldsymbol{y} \in \mathcal{Y}. \end{array}$$

Here the matrix $\boldsymbol{M} = \frac{(a_2 - a_1)^2}{4} \boldsymbol{\Sigma} - a_1 a_2 \boldsymbol{\mu} \boldsymbol{\mu}^T$.

Proof of Theorem 4

Following the proof for Theorem 3, the objective function of Problem (22) can be reformulated as Problem (37), with the constraint equivalent to $s_2 > 0$ together with constraint (38). We note that the constraint (38) can be reformulated as $s_0 \geq \frac{(s_1 + a_n)^2}{4 s_2} - a_n \boldsymbol{y}^T \boldsymbol{\mu} + b_n k,\ n \in \{1, 2\}$. Hence, we have

$$s_0 \geq \max \left\{ \frac{(s_1 + a_1)^2}{4 s_2} - a_1 \boldsymbol{y}^T \boldsymbol{\mu} + b_1 k, \frac{(s_1 + a_2)^2}{4 s_2} - a_2 \boldsymbol{y}^T \boldsymbol{\mu} + b_2 k \right\}.$$

The optimal s_1 solves the following equation,

$$\frac{(s_1+a_1)^2}{4s_2} - a_1\boldsymbol{y}^T\boldsymbol{\mu} + b_1k = \frac{(s_1+a_2)^2}{4s_2} - a_2\boldsymbol{y}^T\boldsymbol{\mu} + b_2k$$

$$s_1 = \frac{1}{2}\left(-a_1 - a_2 + \frac{4s_2}{a_1-a_2}\left((a_1-a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_2-b_1)\,k\right)\right)$$

$$s_0 = \frac{(s_1+a_1)^2}{4s_2} - a_1\boldsymbol{y}^T\boldsymbol{\mu} + b_1k = As_2 + \frac{B}{s_2} + C,$$

where

$$A = \frac{\left((a_1-a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_2-b_1)\,k\right)^2}{(a_1-a_2)^2}$$

$$B = \frac{(a_1-a_2)^2}{16}$$

$$C = \frac{1}{2}\left(-(a_1+a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_1+b_2)\,k\right).$$

Therefore, the optimal value for the problem (37), and hence the objective function of Problem (22) is

$$\min_{s_2>0}\left\{\left(A + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)s_2 + \frac{B}{s_2} + C\right\} = 2\sqrt{\left(A + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)B} + C.$$

The Problem (22) has optimal value no greater than k is equivalent to

$$2\sqrt{\left(A + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)B} + C \le k.$$

Then we have

$$\begin{cases} 4\left(A + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)B \le (C-k)^2 \\ k - C \ge 0. \end{cases} \tag{42}$$

For the first constraint in (42),

$$4\left(A + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)B \le (C-k)^2$$

$$4\left(\frac{\left((a_1-a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_2-b_1)\,k\right)^2}{(a_1-a_2)^2} + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)\frac{(a_1-a_2)^2}{16} \le \frac{1}{4}\left(-(a_1+a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_1+b_2-2)\,k\right)^2$$

$$\left((a_1-a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_2-b_1)\,k\right)^2 + \boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\,(a_1-a_2)^2 \le \left(-(a_1+a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_1+b_2-2)\,k\right)^2$$

$$\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\,(a_1-a_2)^2 \le \left(-2a_2\boldsymbol{y}^T\boldsymbol{\mu} + 2(b_2-1)k\right)\left(-2a_1\boldsymbol{y}^T\boldsymbol{\mu} + 2(b_1-1)k\right)$$

$$\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\,(a_1-a_2)^2 \le 4\left(-a_2\boldsymbol{y}^T\boldsymbol{\mu} + (b_2-1)k\right)\left(-a_1\boldsymbol{y}^T\boldsymbol{\mu} + (b_1-1)k\right).$$

This constraint is indeed

$$\begin{aligned} \frac{(a_2-a_1)^2}{4}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} &\le a_1a_2\left(\boldsymbol{y}^T\boldsymbol{\mu}\right)^2 - (a_1(b_2-1) + a_2(b_1-1))k\boldsymbol{\mu}^T\boldsymbol{y} + (b_1-1)(b_2-1)k^2 \\ &= a_1a_2\boldsymbol{y}^T\boldsymbol{\mu}\boldsymbol{\mu}^T\boldsymbol{y} - a_1(b_2-1)k\boldsymbol{\mu}^T\boldsymbol{y}, \end{aligned}$$

where the last equality holds from $b_1 = 1$. It can be further reformulated as

$$a_1(b_2-1)k\boldsymbol{\mu}^T\boldsymbol{y} + \boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \le 0,$$

where

$$\boldsymbol{M} = \frac{(a_2-a_1)^2}{4}\boldsymbol{\Sigma} - a_1a_2\boldsymbol{\mu}\boldsymbol{\mu}^T$$

is symmetric but not necessarily positive-semidefinite.

For the second constraint in (42), given $b_1 = 1$, we have

$$\frac{1}{2}\Big(-(a_1+a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} + (b_1+b_2)\,k\Big) \le k$$

$$(1-b_2)k + (a_1+a_2)\,\boldsymbol{y}^T\boldsymbol{\mu} \ge 0.$$

Q.E.D.

Proposition 5 *Consider the case that $\mathcal{P} = \mathcal{P}_M$. Then there exists $n^* \in \mathcal{N}$ such that Problem (22) is equivalent to the following problem,*

$$\begin{array}{ll} \min & \dfrac{(s_1+a_{n^*})^2}{4s_2} + b_{n^*}k + s_1\boldsymbol{\mu}^T\boldsymbol{y} + s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \\ \text{s.t.} & 2s_1a_{n^*} + 4b_{n^*}ks_2 + a_{n^*}^2 \ge 2s_1a_n + 4b_nks_2 + a_n^2 \qquad \forall n \in \mathcal{N}\setminus\{n^*\} \\ & s_2 > 0,\;\; s_1 \in \Re,\;\; \boldsymbol{y} \in \mathcal{Y}, \end{array}$$

where $\boldsymbol{M} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T$.

Proof of Proposition 5

Given the information set $\mathcal{P}$ as $\mathcal{P}_M$, we can use the projection theorem by Popescu (2007) and reformulate the value of equation (36) as

$$\sup_{\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]=\boldsymbol{\mu},\; \mathbb{E}_{\mathbb{P}}[(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})(\tilde{\boldsymbol{z}}-\boldsymbol{\mu})^T]=\boldsymbol{\Sigma}} \mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\left\{-a_n\boldsymbol{y}^T\tilde{\boldsymbol{z}} + b_nk\right\}\right] = \sup_{\mathbb{E}_{\mathbb{P}}[\tilde{x}]=\boldsymbol{y}^T\boldsymbol{\mu},\; \mathbb{E}_{\mathbb{P}}[\tilde{x}^2]=\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}} \mathbb{E}_{\mathbb{P}}\left[\max_{n\in\mathcal{N}}\left\{-a_n\tilde{x} + b_nk\right\}\right].$$

where $\boldsymbol{M} = \boldsymbol{\Sigma} + \boldsymbol{\mu}\boldsymbol{\mu}^T$. The dual of the above optimization problem is

$$\begin{array}{ll} \inf & s_0 + s_1\boldsymbol{\mu}^T\boldsymbol{y} + s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \\ \text{s.t.} & s_0 + s_1x + s_2x^2 \ge -a_nx + b_nk, \qquad \forall n \in \mathcal{N}, x \in \Re. \end{array} \tag{43}$$

Note that the constraint being true for all $x \in \Re$ implies feasible s_2 must be strictly positive. In that case, for any $n \in \mathcal{N}$,

$$s_0 + s_1x + s_2x^2 \ge -a_nx + b_nk \qquad \forall x \in \Re$$

is equivalent to

$$s_0 \ge \max_{x\in\Re}\{-s_2x^2 + (-a_n - s_1)x + b_nk\} = \frac{(s_1+a_n)^2}{4s_2} + b_nk.$$

Therefore, Problem (43) is

$$\inf_{s_2>0, s_1\in\Re} \left\{\max_{n\in\mathcal{N}}\left\{\frac{(s_1+a_n)^2}{4s_2} + b_nk\right\} + s_1\boldsymbol{\mu}^T\boldsymbol{y} + s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y}\right\}. \tag{44}$$

Denote the optimal n by n^*. Problem (44) can be equivalently solved by

$$\begin{array}{ll} \min & \dfrac{(s_1+a_{n^*})^2}{4s_2} + b_{n^*}k + s_1\boldsymbol{\mu}^T\boldsymbol{y} + s_2\boldsymbol{y}^T\boldsymbol{M}\boldsymbol{y} \\ \text{s.t.} & \dfrac{(s_1+a_{n^*})^2}{4s_2} + b_{n^*}k \ge \dfrac{(s_1+a_n)^2}{4s_2} + b_nk \qquad \forall n \in \mathcal{N}\setminus\{n^*\} \\ & s_2 > 0,\;\; s_1 \in \Re,\;\; \boldsymbol{y} \in \mathcal{Y}, \end{array}$$

which is indeed Problem (24). Q.E.D.

Lemma 2 *The optimal value of Problem (26) is no greater than k if and only if* $f_1(n^*,n^0) \le 4k^2(1-b_{n^*})(1-b_{n^0})$*, where the function* $f_1(n^*,n^0)$ *is defined as*

$$f_1(n^*,n^0) = \min_{\boldsymbol{y}\in\mathcal{Y}} \left\{ \boldsymbol{y}^T \left((a_{n^o}-a_{n^*})^2 \boldsymbol{\Sigma} - 4a_{n^o}a_{n^*}\boldsymbol{\mu}\boldsymbol{\mu}^T \right) \boldsymbol{y} - 4k\left(a_{n^o}\left(1-b_{n^*}\right)+a_{n^*}\left(1-b_{n^o}\right)\right)\boldsymbol{\mu}^T\boldsymbol{y} \right\}.$$

Proof of Lemma 2

Similar to the proof for Theorem 4, we reformulate the constraint as follows,

$$\frac{1}{2}\sqrt{\left(\left(k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}-\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)(a_{n^o}-a_{n^*})^2}+\frac{1}{2}\left((b_{n^o}+b_{n^*})k-(a_{n^o}+a_{n^*})\boldsymbol{\mu}^T\boldsymbol{y}\right)\le k$$

$$\sqrt{\left(\left(k\frac{b_{n^o}-b_{n^*}}{a_{n^o}-a_{n^*}}-\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\right)(a_{n^o}-a_{n^*})^2}\le(a_{n^o}+a_{n^*})\boldsymbol{\mu}^T\boldsymbol{y}+k\left(2-b_{n^o}-b_{n^*}\right)$$

$$\left(k\left(b_{n^o}-b_{n^*}\right)-\left(a_{n^o}-a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+\left(a_{n^o}-a_{n^*}\right)^2\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\le\left(\left(a_{n^o}+a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}+k\left(2-b_{n^o}-b_{n^*}\right)\right)^2$$

$$\begin{aligned}\left(a_{n^o}-a_{n^*}\right)^2\left(\boldsymbol{\mu}^T\boldsymbol{y}\right)^2-2k\left(b_{n^o}-b_{n^*}\right)\left(a_{n^o}-a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}+k^2\left(b_{n^o}-b_{n^*}\right)^2+\left(a_{n^o}-a_{n^*}\right)^2\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\le&\\ \left(a_{n^o}+a_{n^*}\right)^2\left(\boldsymbol{\mu}^T\boldsymbol{y}\right)^2+2k\left(2-b_{n^o}-b_{n^*}\right)\left(a_{n^o}+a_{n^*}\right)\boldsymbol{\mu}^T\boldsymbol{y}+k^2\left(2-b_{n^o}-b_{n^*}\right)^2&\end{aligned}$$

$$-4a_{n^o}a_{n^*}\left(\boldsymbol{\mu}^T\boldsymbol{y}\right)^2-4k\left(a_{n^o}\left(1-b_{n^*}\right)+a_{n^*}\left(1-b_{n^o}\right)\right)\boldsymbol{\mu}^T\boldsymbol{y}+\left(a_{n^o}-a_{n^*}\right)^2\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y}\le 4k^2\left(1-b_{n^o}\right)\left(1-b_{n^*}\right)$$

$$-4k\left(a_{n^o}\left(1-b_{n^*}\right)+a_{n^*}\left(1-b_{n^o}\right)\right)\boldsymbol{\mu}^T\boldsymbol{y}+\boldsymbol{y}^T\left(\left(a_{n^o}-a_{n^*}\right)^2\boldsymbol{\Sigma}-4a_{n^o}a_{n^*}\boldsymbol{\mu}\boldsymbol{\mu}^T\right)\boldsymbol{y}\le 4k^2\left(1-b_{n^o}\right)\left(1-b_{n^*}\right).$$

Q.E.D.

Theorem 5 *Consider the case that* $v(\boldsymbol{y},\tilde{\boldsymbol{z}})=\boldsymbol{y}^T\tilde{\boldsymbol{z}}$, $\mathcal{P}=\mathcal{P}_M$*. Then Problem (22) has optimal value no greater than k if and only if Enumerative Algorithm, which solves at most* $N(N-1)^2$ *quadratic optimization problem with binary decision variables, returns* $\text{obj}=0$.

Proof of Theorem 5

It is straightforward from the definition of f_1, f_2 and the preceding discussion of the three scenarios for solving Problem (24). For case 2, we enumerate all pairs of $n^*,n^o\in\mathcal{N}$ with $n^*\ne n^o$. There are $N(N-1)$ combinations. For case 3, we choose 3 different n^*,n_1,n_2 from set $\mathcal{N}$ and solve at most $N(N-1)(N-2)$ quadratic problems. Therefore, when implementing Enumerate Algorithm, we solve at most $N(N-1)+N(N-1)(N-2)=N(N-1)^2$ quadratic problems. Q.E.D.

Proposition 6 *For given value of k and* $\beta(\theta)$ *defined by (33), constraints*

$$\text{CVaR}_{1-\beta(\theta)}\left(T_i-\sum_{j=1}^J\tilde{t}_{ij}y_{ij}\right)\le\theta k,\qquad\qquad\forall\theta_i>0, i=1,\dots,I$$

can be formulated as a set of quadratic constraints.

Proof of Proposition 6

Observe that by definition, for any random variable $\tilde{x}$ and $\alpha\in(0,1)$,

$$\text{CVaR}_\alpha(\tilde{x})=\inf_\nu\left\{\nu+\frac{1}{1-\alpha}\sup_{\mathbb{P}\in\mathcal{P}}\mathbb{E}_{\mathbb{P}}\left[(-\nu-\tilde{x})^+\right]\right\}.$$

Then we reformulate the constraint as follows,

$$\begin{aligned}
& \mathrm{CVaR}_{1-\beta(\theta)}(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}}) \le \theta k,\ \forall \theta > 0 \\
\Leftrightarrow\ & \sup_{\theta>0} \left\{ \mathrm{CVaR}_{1-\beta(\theta)}(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}}) - \theta k \right\} \le 0 \\
\Leftrightarrow\ & \sup_{\theta>0}\inf_{\nu} \left\{ \nu + (\theta+1)\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] - \theta k \right\} \le 0 \\
\Leftrightarrow\ & \inf_{\nu}\sup_{\theta>0} \left\{ \nu + (\theta+1)\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] - \theta k \right\} \le 0 \\
\Leftrightarrow\ & \sup_{\theta>0} \left\{ (\theta+1)\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] - (\theta k - \nu) \right\} \le 0 \\
\Leftrightarrow\ & \sup_{\theta>0} \left\{ \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] - \frac{\theta k - \nu}{\theta+1} \right\} \le 0 \\
\Leftrightarrow\ & \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] \le \inf_{\theta>0} \frac{\theta k - \nu}{\theta+1}.
\end{aligned} \tag{45}$$

Here the third equivalence holds since the function inside the big bracket is concave in θ and convex in ν and hence by Sion's minimax Theorem (Sion 1958), $\sup_{\theta>0}$ and $\inf_\nu$ can be exchanged.

We next see the problem $\inf_{\theta>0} f(\theta)$ where $f(\theta) := \frac{\theta k-\nu}{\theta+1}$. Taking derivative, we have $f'(\theta) = \frac{k+\nu}{(\theta+1)^2}$. We consider two cases.

- Case 1: if $k+\nu \ge 0$, i.e., $k \ge -\nu$, then $f'(\theta) \ge 0$ and $f(\theta)$ is non-decreasing in θ. We have $\inf_{\theta>0} f(\theta) = f(0) = -\nu$.
- Case 2: if $k+\nu < 0$, i.e., $k < -\nu$, then $f'(\theta) < 0$ and $f(\theta)$ is decreasing in θ. We have $\inf_{\theta>0} f(\theta) = \lim_{\theta\to\infty} f(\theta) = k$.

Combining two cases, we have $\inf_{\theta>0} f(\theta) = \min\{k, -\nu\}$. Therefore, the constraint in (45) is equivalent to

$$\begin{cases} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] \le k \\ \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}}\left[\left(-\nu-(T-\boldsymbol{y}^T\tilde{\boldsymbol{t}})\right)^+ \right] \le -\nu. \end{cases} \tag{46}$$

Following the same analysis of Theorem 4, $\sup_{\mathbb{P}\in\mathcal{P}_M} \mathbb{E}_{\mathbb{P}}\left[\max\{-\nu-(T-\tilde{\boldsymbol{t}}^T\boldsymbol{y}), 0\} \right] \le k$ is equivalent to

$$\begin{cases} k(\boldsymbol{\mu}^T\boldsymbol{y} - T - \nu - k) + \frac{1}{4}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} \le 0 \\ k + \frac{1}{2}(\nu + T - \boldsymbol{\mu}^T\boldsymbol{y}) \ge 0, \end{cases}$$

and $\sup_{\mathbb{P}\in\mathcal{P}_M} \mathbb{E}_{\mathbb{P}}\left[\max\{-\nu-(T-\tilde{\boldsymbol{t}}^T\boldsymbol{y}), 0\} \right] \le -\nu$ is equivalent to

$$\begin{cases} \nu(T - \boldsymbol{\mu}^T\boldsymbol{y}) + \frac{1}{4}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} \le 0 \\ -\nu + \frac{1}{2}(\nu + T - \boldsymbol{\mu}^T\boldsymbol{y}) \ge 0. \end{cases}$$

However, there's a product of two decision variable $\nu\boldsymbol{y}$. We use big-M method to replace $\nu\boldsymbol{y}$ with $\boldsymbol{p}-\boldsymbol{q}$ and add constraints

$$\begin{array}{ll}
\nu = \hat{p} - \hat{q} & \\
\boldsymbol{p} \le \hat{p}\mathbf{1} & \boldsymbol{q} \le \hat{q}\mathbf{1} \\
\boldsymbol{p} \le M\boldsymbol{y} & \boldsymbol{q} \le M\boldsymbol{y} \\
\boldsymbol{p} \ge \hat{p}\mathbf{1} - M(\mathbf{1}-\boldsymbol{y}) & \boldsymbol{q} \ge \hat{q}\mathbf{1} - M(\mathbf{1}-\boldsymbol{y}) \\
\boldsymbol{p} \ge \mathbf{0} & \boldsymbol{q} \ge \mathbf{0} \\
\hat{p} \ge 0 & \hat{q} \ge 0,
\end{array}$$

where $\mathbf{1}$ is the vector with all entries being 1. Then the constraints (46) are equivalent to

$$\begin{cases} k(\boldsymbol{\mu}^T\boldsymbol{y} - T - \nu - k) + \frac{1}{4}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} \le 0 \\ k + \frac{1}{2}(\nu + T - \boldsymbol{\mu}^T\boldsymbol{y}) \ge 0 \\ \nu T - \boldsymbol{\mu}^T(\boldsymbol{p} - \boldsymbol{q}) + \frac{1}{4}\boldsymbol{y}^T\boldsymbol{\Sigma}\boldsymbol{y} \le 0 \\ -\nu + \frac{1}{2}(\nu + T - \boldsymbol{\mu}^T\boldsymbol{y}) \ge 0 \\ \nu = \hat{p} - \hat{q} \\ \boldsymbol{p} \le \hat{p}\mathbf{1} \\ \boldsymbol{q} \le \hat{q}\mathbf{1} \\ \boldsymbol{p} \le M\boldsymbol{y} \\ \boldsymbol{q} \le M\boldsymbol{y} \\ \boldsymbol{p} \ge \hat{p}\mathbf{1} - M(\mathbf{1} - \boldsymbol{y}) \\ \boldsymbol{q} \ge \hat{q}\mathbf{1} - M(\mathbf{1} - \boldsymbol{y}) \\ \boldsymbol{p}, \boldsymbol{q} \ge \mathbf{0} \\ \hat{p}, \hat{q} \ge 0. \end{cases}$$

Q.E.D.

Appendix B: Mathematical Derivation for Examples 1 and 3

Example 1. Choose the underlying monetary risk measure as $\mu_\theta(\tilde{x}) = \text{VaR}_{1-\beta(\theta)}$, where $\beta : \Re_+ \to (0,1)$ is a non-increasing function. The corresponding REM constructed by Equation (5) is

$$\rho_{\text{VaR},\beta}(\tilde{x}) = \inf\{k > 0 \ : \ \text{VaR}_{1-\beta(\theta)}(\tilde{x}) \le \theta k, \ \forall \theta > 0\} = \inf\{k > 0 \ : \ \mathbb{P}(-\tilde{x} > k\theta) \le \beta(\theta), \ \forall \theta > 0\},$$

which is indeed the PEM defined in Definition 1, i.e., $\rho_{\text{VaR},\beta}(\tilde{x}) = \text{PEM}_\beta(\tilde{x})$.

Example 3. Choose the underlying monetary risk measure as

$$\mu_\theta(\tilde{x}) = \inf_{\alpha>0} \left\{ \alpha \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha}\right)\right] + \alpha\theta \right\},$$

which is the coherent version of the entropic risk measure $\theta \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\theta}\right)\right]$. The corresponding CREM constructed according to Equation (6) is

$$\rho_{\text{Entropic}}(\tilde{x}) = \inf\left\{ k > 0 \ : \ \inf_{\alpha>0}\left\{ \alpha \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha}\right)\right] + \alpha\theta \right\} \le \theta k, \ \forall \theta > 0 \right\}. \tag{47}$$

We now show that the representation of $\rho_{\text{Entropic}}(\tilde{x})$ can indeed be simplified as

$$\rho_{\text{Entropic}}(\tilde{x}) = \inf\left\{ k > 0 \ : \ k \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{k}\right)\right] \le 0 \right\} \tag{48}$$

by demonstrating the equivalence between the set in the RHS of Equation (47) and that in the RHS of Equation (48). To this end, we consider any $k > 0$. If $k \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{k}\right)\right] \le 0$, then $\forall \theta > 0$, and we have

$$\inf_{\alpha>0}\left\{ \alpha \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha}\right)\right] + \alpha\theta \right\} \le k \ln \mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{k}\right)\right] + k\theta \le \theta k.$$

The case of $k\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{k}\right)\right]>0$ can be proved as follows. We observe that $\alpha\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha}\right)\right]$ is 1) non-increasing in $\alpha>0$ (see, for instance, Hall et al. 2015), and 2) convex in $\alpha>0$ since $\forall\lambda\in(0,1),\alpha_1,\alpha_2>0$,

$$\begin{aligned}
&\lambda\alpha_1\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha_1}\right)\right]+(1-\lambda)\alpha_2\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha_2}\right)\right]\\
&=\alpha_\lambda\ln\left(\left(\mathbb{E}\left[\left(\exp\left(-\frac{\lambda\tilde{x}}{\alpha_\lambda}\right)\right)^{\frac{\alpha_\lambda}{\lambda\alpha_1}}\right]\right)^{\frac{\lambda\alpha_1}{\alpha_\lambda}}\times\left(\mathbb{E}\left[\left(\exp\left(-\frac{(1-\lambda)\tilde{x}}{\alpha_\lambda}\right)\right)^{\frac{\alpha_\lambda}{(1-\lambda)\alpha_2}}\right]\right)^{\frac{(1-\lambda)\alpha_2}{\alpha_\lambda}}\right)\\
&\geq\alpha_\lambda\ln\mathbb{E}\left[\exp\left(-\frac{\lambda\tilde{x}}{\alpha_\lambda}\right)\exp\left(-\frac{(1-\lambda)\tilde{x}}{\alpha_\lambda}\right)\right]\\
&=\alpha_\lambda\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha_\lambda}\right)\right],
\end{aligned}$$

where we denote $\alpha_\lambda=\lambda\alpha_1+(1-\lambda)\alpha_2$, and the inequality follows from Hölder's Inequality. Thus, we can always choose $\theta^*>0$ such that 0 is in the subdifferential of $\left(\alpha\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha}\right)\right]+\alpha\theta^*\right)$ at $\alpha=k$, then we have

$$\inf_{\alpha>0}\left\{\alpha\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{\alpha}\right)\right]+\alpha\theta^*\right\}=k\ln\mathbb{E}\left[\exp\left(-\frac{\tilde{x}}{k}\right)\right]+k\theta^*>k\theta^*.$$

Hence, the set in the RHS of Equation (47) equals to that in the RHS of Equation (48), and the representation of $\rho_{\text{Entropic}}(\tilde{x})$ in (48) is equivalent to the original definition in (47).

References

Ahmed MA, Alkhamis TM (2009) Simulation optimization for an emergency department healthcare unit in kuwait. *European Journal of Operational Research* 198(3):936 – 942, ISSN 0377-2217.

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

Aumann RJ, Serrano R (2008) An economic index of riskiness. *Journal of Political Economy* 116(5):810–836.

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

Chen X, Sim M, Simchi-Levi D, Sun P (2007) Risk aversion in inventory management. *Operations Research* 55(5):828–842.

Chen YF, Xu M, Zhang ZG (2009) A risk-averse newsvendor model under the cvar criterion. *Operations Research* 57(4):1040–1044.

Choi S, Ruszczyński A (2008) A risk-averse newsvendor with law invariant coherent measures of risk. *Operations Research Letters* 36(1):77–82.

Choi S, Ruszczynski A, Zhao Y (2011) A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research* 59(2):346–364.

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

Deng Y, Shen S (2016) Decomposition algorithms for optimizing multi-server appointment scheduling with chance constraints. *Mathematical Programming* 157(1):245–276.

Denton BT, Miller AJ, Balasubramanian HJ, Huschka TR (2010) Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research* 58(4-part-1):802–816.

Eeckhoudt L, Gollier C, Schlesinger H (1995) The risk-averse (and prudent) newsboy. *Management Science* 41(5):786–794.

Evans R, Cantoni A, Fortmann T (1977a) Envelope-constrained filters–ii: Adaptive structures. *IEEE Transactions on Information Theory* 23(4):435–444.

Evans R, Fortmann T, Cantoni A (1977b) Envelope-constrained filters–i: Theory and applications. *IEEE Transactions on Information Theory* 23(4):421–434.

Föllmer H, Schied A, Lyons TJ (2004) *Stochastic Finance: An Introduction in Discrete Time* (Walter de Gruyter, Berlin, Germany).

Hall NG, Long DZ, Qi J, Sim M (2015) Managing underperformance risk in project portfolio selection. *Operations Research* 63(3):660–675.

Hanasusanto GA, Roitch V, Kuhn D, Wiesemann W (2017) Ambiguous joint chance constraints under mean and dispersion information. *Operations Research* 65(3):751–767.

Jorion P (2006) *Value at Risk: the New Benchmark for Controlling Market Risk* (McGraw-Hill New York), 3 edition.

Levy H (1992) Stochastic dominance and expected utility: Survey and analysis. *Management Science* 38(4):555–593.

Nemirovski A (2012) On safe tractable approximations of chance constraints. *European Journal of Operational Research* 219(3):707–718.

Payne JW, Laughhunn DJ, Crum R (1980) Translation of gambles and aspiration level effects in risky choice behavior. *Management Science* 26(10):1039–1060.

Payne JW, Laughhunn DJ, Crum R (1981) Further tests of aspiration level effects in risky choice behavior. *Management Science* 27(8):953–958.

Pender J (2016) Risk measures and their application to staffing nonstationary service systems. *European Journal of Operational Research* 254(1):113 – 126, ISSN 0377-2217.

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

Rockafellar RT, Uryasev S (2000) Optimization of conditional value-at-risk. *Journal of Risk* 2:21–42.

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

Rodriguez-Alvarez A, Roibás D, Wall A (2012) Reserve capacity of public and private hospitals in response to demand uncertainty. *Health Economics* 21(7):839–851.

Scarf H (1958) A min-max solution of an inventory problem. *Studies in the mathematical theory of inventory and production* 10:201–209.

Shylo OV, Prokopyev OA, Schaefer AJ (2013) Stochastic operating room scheduling for high-volume specialties under block booking. *INFORMS Journal on Computing* 25(4):682–692.

Sion M (1958) On general minimax theorems. *Pacific Journal of Mathematics* 8(1):171–176.

Wiesemann W, Kuhn D, Sim M (2014) Distributionally robust convex optimization. *Operations Research* 62(6):1358–1376.

Xie W, Ahmed S (2018) On deterministic reformulations of distributionally robust joint chance constrained optimization problems. *SIAM Journal on Optimization* 28(2):1151–1182.

Xu H, Caramanis C, Mannor S (2012) Optimization under probabilistic envelope constraints. *Operations Research* 60(3):682–699.

Zhang Y, Jiang R, Shen S (2018) Ambiguous chance-constrained binary programs under mean-covariance information. *SIAM Journal on Optimization* 28(4):2922–2944.