

# Tight Approximations of Dynamic Risk Measures

Dan A. Iancu\*

Graduate School of Business, Stanford University, Stanford, CA 94305, [daniancu@stanford.edu](mailto:daniancu@stanford.edu)

Marek Petrik

IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, [mpetrik@us.ibm.com](mailto:mpetrik@us.ibm.com)

Dharmashankar Subramanian

IBM T.J. Watson Research Center, Yorktown Heights, NY 10598, [dharmash@us.ibm.com](mailto:dharmash@us.ibm.com)

This paper compares two different frameworks recently introduced in the literature for measuring risk in a multi-period setting. The first corresponds to applying a single coherent risk measure to the cumulative future costs, while the second involves applying a composition of one-step coherent risk mappings. We summarize the relative strengths of the two methods, characterize several necessary and sufficient conditions under which one of the measurements always dominates the other, and introduce a metric to quantify how close the two risk measures are.

Using this notion, we address the question of how tightly a given coherent measure can be approximated by lower or upper-bounding compositional measures. We exhibit an interesting asymmetry between the two cases: the tightest possible upper-bound can be exactly characterized, and corresponds to a popular construction in the literature, while the tightest-possible lower bound is not readily available. We show that testing domination and computing the approximation factors is generally NP-hard, even when the risk measures in question are comonotonic and law-invariant. However, we characterize conditions and discuss several examples where polynomial-time algorithms are possible. One such case is the well-known Conditional Value-at-Risk measure, which is further explored in our companion paper Huang et al. [32].

Our theoretical and algorithmic constructions exploit interesting connections between the study of risk measures and the theory of submodularity and combinatorial optimization, which may be of independent interest.

*Key words:* coherent risk measures, dynamic consistency, tight approximations, submodular functions, polymatroids

---

**1. Introduction.** Measuring the intrinsic risk in a particular unknown outcome and comparing multiple risky alternatives has been a topic of central concern in a wide range of academic disciplines, resulting in the development of numerous frameworks, such as expected utility, stochastic ordering, and, in recent years, convex and coherent risk measures.

The latter class has emerged as an axiomatically justified and computationally tractable alternative to several classical approaches, and has provided a strong bridge across a variety of parallel streams of research, including ambiguous representations of preferences in economics (e.g., Gilboa and Schmeidler [28], Schmeidler [51], Epstein and Schneider [22], Maccheroni et al. [37]), axiomatic treatments of market risk in financial mathematics (Artzner et al. [5], Föllmer and Schied [24]), actuarial science (Wirch and Hardy [58], Wang [57], Acerbi [2], Kusuoka [36], Tsanakas [56]), operations research (Ben-Tal and Teboulle [11]) and statistics (Huber [33]). As such, our goal in the present paper is not to *motivate* the use of risk measures – rather, we take the framework as given, and investigate two distinct ways of using it to ascribe risk in dynamic decision settings.

A first approach, prevalent among practitioners, entails applying a static risk measure to the total future costs accumulated over the remaining problem horizon, and conditioned on the available information. More formally, a decision maker faced with a future sequence of random costs

\*Part of this research was conducted while the author was with the Department of Mathematical Sciences of the IBM T.J. Watson Research Center, whose financial support is gratefully acknowledged.

$X_t, \dots, X_T$ , respectively dispensed over a finite horizon  $t, t+1, \dots, T$ , would measure the risk at time  $t$  by  $\mu_t(X_t + \dots + X_T | \mathcal{F}_t)$ , where  $\mathcal{F}_t$  denotes the filtration containing all information at time  $t$ , and  $\mu_t$  is a static risk measure. In practice, the same  $\mu_t = \mu$  is often used at every time  $t$ , resulting in a risk preference that is easy to specify and calibrate. Apart from simplicity, the approach also has one other key advantage: when the risk measure used is convex, *static* decisions can be efficiently computed by combining simulation procedures with convex optimization (e.g., Rockafellar and Uryasev [43], Ruszczyński and Shapiro [49]). This has led to a wide adoption of the methodology in practice, as well as in several academic papers (see, e.g., Basak and Shapiro [9], Cuoco et al. [20] and references therein).

The paradigm above, however, is known to suffer from several serious shortcomings. It can result in inconsistent preferences over risk profiles in time, whereby a decision maker faced with two alternative cumulative costs  $Y$  and  $Z$  can deem  $Y$  riskier than  $Z$  in every state of the world at some time  $t+1$ , but nonetheless deem  $Z$  riskier than  $Y$  at time  $t$ . This *dynamic* or *time inconsistency* has been criticized from an axiomatic perspective, as it is a staple of irrational behavior [22, 45, 6]. Furthermore, time inconsistent objectives couple risk preferences over time, which is very undesirable from a dynamic optimization viewpoint, since it prevents applying the principles of Dynamic Programming to decompose the problem in stages (Epstein and Schneider [22], Ruszczyński and Shapiro [48], Nilim and El Ghaoui [38], Iyengar [34]).

In order to correct such undesirable effects, additional conditions must be imposed on the risk measurement process at distinct time periods. Such requirements have been discussed extensively in the literature, and it has been shown that any risk measure that is *time consistent* is obtained by composing one-step conditional risk mappings. More formally, a time consistent decision maker faced with costs  $X_1, \dots, X_T$  would assess the risk at time  $t$  by  $\mu_t(\mu_{t+1}(\dots \mu_T(X_t + \dots + X_T | \mathcal{F}_t) \dots))$ , for a set of suitable mappings  $\{\mu_\tau\}_{\tau \in \{t, \dots, T\}}$  (see, e.g., Epstein and Schneider [22], Riedel [40], Cheridito et al. [15], Artzner et al. [6], Roorda et al. [45], Föllmer and Penner [23], Ruszczyński [47]). Apart from yielding consistent preferences, this compositional form also allows a recursive estimation of the risk, and an application of the Bellman principle in optimization problems involving dynamic risk measures [38, 34, 48].

From a pragmatic perspective, however, the compositional form entails a significantly more complicated risk assessment than the naïve inconsistent approach. A risk manager would need to specify single-period conditional risk mappings for every future time-point; furthermore, even if these corresponded to the same risk measure  $\mu$ , the exact result of the composition would no longer be easily interpretable, and would bear no immediate relation to the original  $\mu$ . Our conversations with managers also revealed a certain feeling that such a measurement could result in “overly conservative” assessments, since risks are compounded in time – for instance, by composing VaR, one would obtain extreme quantiles of quantities that are already extreme quantiles. This has been recognized informally in the literature by Roorda and Schumacher [46, 44], who proposed new notions of time consistency that avoided the issue, but without establishing formally if or to what degree the conservatism is actually true. Furthermore, it is not obvious how “close” a particular compositional measure is to a given inconsistent one, and how one could go about constructing the latter in a way that tightly approximates the former. This issue should be very relevant when considering dynamic decision problems under risk, but it seems to have been largely ignored by the literature (most papers examining operational problems under dynamic risk typically start with a set of given dynamic risk measures, e.g., Ahmed et al. [4], Shapiro [54], Choi et al. [17]).

With this motivation in mind, the goal of the present paper is to better understand the relation and exact tradeoffs between the two measurement processes outlined above, and to provide guidelines for constructing and/or estimating safe counterparts of one from the other. Our contributions are as follows.

- We provide several equivalent necessary and sufficient conditions guaranteeing when a time consistent risk measure  $\mu_C$  always over (or under) estimates risk as compared with an inconsistent measure  $\mu_I$ . We argue that iterating the same  $\mu_I$  does not necessarily over (or under) estimate risk as compared to a single static application of  $\mu_I$ , and this is true even in the case considered by Roorda and Schumacher [46, 44]. We show that composition with conditional expectation operators at *any* stage of the measurement process results in valid, time consistent lower bounds. By contrast, upper bounds are obtained only when composing with worst-case operators in the *last* stage of the measurement process.

- We formalize the problem of characterizing and computing the smallest  $\alpha_{\mu_C, \mu_I}$  and  $\alpha_{\mu_I, \mu_C}$  such that  $\mu_C \leq \mu_I \leq \alpha_{\mu_C, \mu_I} \cdot \mu_C$  and  $\mu_I \leq \mu_C \leq \alpha_{\mu_I, \mu_C}$ , respectively. The smallest such factors,  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$ , provide a compact notion of how closely a given  $\mu_I$  can be multiplicatively approximated through lower (respectively, upper) bounding consistent measures  $\mu_C$ , respectively. Since, in practice,  $\mu_I$  may be far easier to elicit from observed preferences or to estimate from empirical data, characterizing and computing  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  can be seen as the first step towards *constructing* the time-consistent risk measure  $\mu_C$  that is “closest” to a given  $\mu_I$ .

- Using results from the theory of submodularity and matroids, we particularize our results to the case when  $\mu_I$  and  $\mu_C$  are both comonotonic risk measures. We show that computing  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  is generally NP-hard, even when the risk measures in question are law-invariant. However, we provide several conditions under which the computation becomes simpler. Using these results, we compare the strength of approximating a given  $\mu_I$  by time-consistent measures obtained through composition with conditional expectation or worst-case operators.

- We characterize the tightest possible time-consistent and coherent upper bound for a given  $\mu_I$ , and show that it corresponds to a construction suggested in several papers in the literature [22, 45, 6, 54], which involves “rectangularizing” the set of probability measures corresponding to  $\mu_I$ . This yields not only the smallest possible  $\alpha_{\mu_I, \mu_C}^*$ , but also the uniformly tightest upper bound among all coherent upper bounds.

- We summarize results from our companion paper [32], which applies the ideas derived here to the specific case when both  $\mu_I$  and  $\mu_C$  are given by Average Value at Risk, a popular measure in financial mathematics. In this case, the results take a considerably simpler form: analytical expressions are available for two-period problems, and polynomial-time algorithms are available for some multi-period problems. We give an exact analytical characterization for the tightest uniform upper bound to  $\mu_I$ , and show that it corresponds to a compositional AVaR risk measure that is increasingly conservative in time. For the case of lower bounds, we give an analytical characterization for two-period problems. Interestingly, we find that the best lower-bounds always provide tighter approximations than the best upper bounds in two-period models, but are also considerably harder to compute than the latter in multi-period models.

The rest of the paper is organized as follows. Section 2 provides the necessary background in static and dynamic risk measures, and introduces the precise mathematical formulation for the questions addressed in the paper. Section 3 discusses the case of determining upper or lower bounding relations between two arbitrary consistent and inconsistent risk measures, and characterizes the resulting factors  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$ . Section 4 discussed our results in detail, touching on the computational complexity, and introducing several examples of how the methodology can be used in practice. Section 5 concludes the paper and suggests directions for future research.

**1.1. Notation.** With  $i < j$ , we use  $[i, j]$  to denote the index set  $\{i, \dots, j\}$ . For a vector  $\mathbf{x} \in \mathbb{R}^n$  and  $i \in \{1, \dots, n\}$ , we use  $x_i$  to denote the  $i$ -th component of  $\mathbf{x}$ . For a set  $S \subseteq \{1, \dots, n\}$ , we let  $\mathbf{x}(S) \stackrel{\text{def}}{=} \sum_{i \in S} x_i$ . Also, we use  $\mathbf{x}_S \in \mathbb{R}^n$  to denote the vector with components  $x_i$  for  $i \in S$  and 0 otherwise (e.g.,  $\mathbf{1}_S$  is the characteristic vector of the set  $S$ ), and  $\mathbf{x}|_S \in \mathbb{R}^{|S|}$  to denote the projection of the vector  $\mathbf{x}$  on the coordinates  $i \in S$ . When no confusion can arise, we denote by  $\mathbf{1}$  the vector

with all components equal to 1. We use  $\mathbf{x}^T$  for the transpose of  $\mathbf{x}$ , and  $\mathbf{x}^T \mathbf{y} \stackrel{\text{def}}{=} \sum_{i=1}^n x_i y_i$  for the scalar product in  $\mathbb{R}^n$ .

For a set or an array  $S$ , we denote by  $\Pi(S)$  the set of all permutations on the elements of  $S$ .  $\pi(S)$  or  $\sigma(S)$  designate one particular such permutation, with  $\pi(i)$  denoting the element of  $S$  appearing in the  $i$ -th position under permutation  $\pi$ .

We use  $\Delta^n$  to denote the probability simplex in  $\mathbb{R}^n$ , i.e.,  $\Delta^n \stackrel{\text{def}}{=} \{\mathbf{p} \in \mathbb{R}_+^n : \mathbf{1}^T \mathbf{p} = 1\}$ . For a set  $P \subseteq \mathbb{R}^n$ , we use  $\text{ext}(P)$  to denote the set of its extreme points.

Throughout the exposition, we adopt the convention that  $\frac{0}{0} = 0$ .

**2. Consistent and Inconsistent Risk Measures.** As discussed in the introduction, the goal of the present paper is to analyze two paradigms for assessing risk in a dynamic setting: a “naïve” one, obtained by applying a static risk measure to the conditional cumulative future costs, and a “sophisticated”, time-consistent method, obtained by composing one-period risk mappings.

In the present section, we briefly review the relevant background material in risk theory, describe the two approaches formally, and then introduce the main questions addressed in the paper.

**2.1. Probabilistic Model.** We begin by describing the probabilistic model. Our notation and framework are closely in line with that of [55], to which we direct the reader for more details.

For simplicity, we consider a scenario tree representation of the uncertainty space, where  $t \in [0, T]$  denotes the time,  $\Omega_t$  is the set of nodes at stage  $t \in [0, T]$ , and  $\mathcal{C}_i$  is the set of children<sup>1</sup> of node  $i \in \Omega_t$ . We also use  $\mathcal{D}_i$  to denote the set of all leaves descending from node  $i$ , i.e., with  $\mathcal{D}_i = \{i\}$ ,  $\forall i \in \Omega_T$ , we recursively define  $\mathcal{D}_i \stackrel{\text{def}}{=} \cup_{j \in \mathcal{C}_i} \mathcal{D}_j$ ,  $\forall i \in \cup_{t=0}^{T-1} \Omega_t$ . Similarly, we define  $\mathcal{D}_U \stackrel{\text{def}}{=} \cup_{i \in U} \mathcal{D}_i$  for any set  $U \subseteq \Omega_t$ .

With the set  $\Omega_T$  of elementary outcomes, we associate the  $\sigma$ -algebra  $\mathcal{F}_T = 2^{\Omega_T}$  of all its subsets, and we consider the filtration  $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \dots \subseteq \mathcal{F}_T$ , where  $\mathcal{F}_t$  is the sub-algebra of  $\mathcal{F}_{t+1}$  that is generated by the sets  $\{\mathcal{C}_i\}_{i \in \Omega_t}$ , for any  $t \in [0, T-1]$ .

We construct a probability space  $(\Omega_T, \mathcal{F}_T, \mathbb{P})$  by introducing a *reference measure*  $\mathbb{P} \in \Delta^{|\Omega_T|}$ , assumed to satisfy<sup>2</sup>  $\mathbb{P} > 0$ . On the space  $(\Omega_T, \mathcal{F}_T, \mathbb{P})$ , we use  $\mathcal{X}_T$  to denote the space of all functions  $X_T : \Omega_T \rightarrow \mathbb{R}$  that are  $\mathcal{F}_T$ -measurable. Since  $\mathcal{X}_T$  is isomorphic with  $\mathbb{R}^{|\Omega_T|}$ , we denote by  $X_T$  the random variable, and by  $\mathbf{X}_T$  the vector in  $\mathbb{R}^{|\Omega_T|}$  of induced scenario-values, and we identify the expectation of  $X_T$  with respect to a measure  $\mathbf{q} \in \Delta^{|\Omega_T|}$  as the scalar product  $\mathbf{q}^T \mathbf{X}_T$ . In a similar fashion, we introduce the sequence  $\mathcal{X}_t$ ,  $t \in [0, T-1]$ , where  $\mathcal{X}_t$  is the sub-space of  $\mathcal{X}_T$  containing functions which are  $\mathcal{F}_t$ -measurable. Note that any function  $X_t \in \mathcal{X}_t$  is constant on every set  $\mathcal{C}_i$ ,  $i \in \Omega_t$ , so that  $X_t$  can also be identified with the vector  $\mathbf{X}_t \in \mathbb{R}^{|\Omega_t|}$ . To this end, in order to simplify the notation, we identify any function  $f : \mathcal{X}_{t+1} \rightarrow \mathcal{X}_t$  with a set of  $|\Omega_t|$  functions, and we write  $f \equiv (f_i)_{i \in \Omega_t}$ , where  $f_i : \mathbb{R}^{|\Omega_{t+1}|} \rightarrow \mathbb{R}$ . Furthermore, since all the functions of this form that we consider correspond to conditional evaluations on the nodes of the tree, we slightly abuse the notation and write  $f \equiv (f_i)_{i \in \Omega_t}$ , where  $f_i : \mathbb{R}^{|\mathcal{C}_i|} \rightarrow \mathbb{R}$ .

**2.2. Static Risk Measures.** Consider a discrete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and let  $\mathcal{X}$  be a linear space of random variables on  $\Omega$ . In this setup, we are interested in appropriate ways of assessing the riskiness of a random *cost* (or loss)  $X \in \mathcal{X}$ . The standard approach in the literature [5, 25] is to use a functional  $\mu : \mathcal{X} \rightarrow \mathbb{R}$  such that  $\mu(X)$  represents the minimal reduction making a cost  $X$  acceptable to the risk manager. The following axiomatic requirements are typically imposed.

**[P1] Monotonicity.** For any  $X, Y \in \mathcal{X}$  such that  $X \geq Y$ ,  $\mu(X) \geq \mu(Y)$ .

**[P2] Translation invariance.** For any  $X \in \mathcal{X}$  and any  $m \in \mathbb{R}$ ,  $\mu(X + m) = \mu(X) + m$ .

<sup>1</sup> In other words,  $\{\mathcal{C}_i, i \in \Omega_t\}$  is a partition of the nodes in  $\Omega_{t+1}$ ,  $\forall t \in \{0, \dots, T-1\}$ .

<sup>2</sup> This is without loss of generality - otherwise, all arguments can be repeated on a tree where leaves with zero probability are removed.

**[P3]** *Convexity.* For any  $X, Y \in \mathcal{X}$ , and any  $\lambda \in [0, 1]$ ,  $\mu(\lambda X + (1 - \lambda)Y) \leq \lambda \mu(X) + (1 - \lambda) \mu(Y)$ .

**[P4]** *Positive homogeneity.* For any  $X \in \mathcal{X}$ , and any  $\lambda \geq 0$ ,  $\mu(\lambda X) = \lambda \mu(X)$ .

**[P5]** *Comonotonicity.*  $\mu(X + Y) = \mu(X) + \mu(Y)$  for any  $X, Y \in \mathcal{X}$  that are comonotone, i.e.,  $[X(\omega) - X(\omega')][Y(\omega) - Y(\omega')] \geq 0$ , for any  $\omega, \omega' \in \Omega$ .

**[P6]** *Law-invariance.*  $\mu(X) = \mu(Y)$  for any  $X, Y \in \mathcal{X}$  such that  $F_X(\cdot) = F_Y(\cdot)$ .

*Monotonicity* requires that a larger cost should always be deemed riskier. *Translation* (or *cash invariance*) gives  $\mu$  an interpretation as capital requirement: typically, a cost  $X$  is deemed acceptable if  $\mu(X) \leq 0$ , so cash invariance implies that  $\mu(X - \mu(X)) = 0$ , i.e.,  $\mu(X)$  is the smallest amount of cost reduction making  $X$  acceptable. *Convexity* suggests that diversification of costs should never increase the risk (or, conversely, that a convex combination of two acceptable costs  $X$  and  $Y$  should also be acceptable), while *positive homogeneity* implies that risk should scale linearly with the size of the cost. *Comonotonicity* implies that the risk in costs that move together (i.e., are comonotone) cannot be diversified by mixtures, while *law-invariance* requires the risk measures to only depend on the probability distribution of the random costs. For an in-depth discussion and critique of these axioms, we direct the reader to [5, 25] and references therein.

Following the common terminology in the literature, we call any functional satisfying [P1-2] a *risk measure*. Any *risk measure* satisfying [P3] is said to be *convex*, and any *convex risk measure* that satisfies [P4] is said to be *coherent*. The main focus of the present paper are functionals that satisfy<sup>3</sup> [P1-5], which are called *comonotonic risk measures*. Some of our results take a simpler form when further restricting attention to the class of *distortion risk measures*, which are all comonotonic risk measures additionally satisfying [P6]. Such measures have been examined in economics, actuarial science, and financial mathematics, and form a well-established class of risk metrics (see, e.g., [50, 57, 56, 19, 36, 2, 3, 25] for more references and details).

One of the main results in the literature is a universal representation theorem for any coherent risk measure, which takes a specialized form in the comonotonic case [50, 25].

**THEOREM 1.** *A risk measure  $\mu$  is coherent if and only if it can be represented as*

$$\mu(X) = \max_{\mathbb{Q} \in \mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X], \quad (1)$$

for some  $\mathcal{Q} \subseteq \Delta^{|\Omega|}$ . Furthermore, if  $\mu$  is comonotonic, then  $\mathcal{Q} = \{\mathbb{Q} \in \Delta^{|\Omega|} : \mathbb{Q}(S) \leq c(S), \forall S \in \mathcal{F}\}$ , where  $c$  is a *Choquet capacity*.

The result essentially states that any coherent risk measure is an expectation with respect to a worst-case probability measure, chosen adversarially from a suitable set of test measures (or generalized scenarios)  $\mathcal{Q}$ . For comonotonic risk measures, this set is uniquely determined by a particular function  $c$ , known as a *Choquet capacity*.

**DEFINITION 1.** A set function  $c : 2^\Omega \rightarrow [0, 1]$  is said to be a *Choquet capacity* if it satisfies the following properties:

- nondecreasing:  $c(A) \leq c(B)$ ,  $\forall A \subseteq B \subseteq \Omega$
- normalized:  $c(\emptyset) = 0$  and  $c(\Omega) = 1$
- submodular:  $c(A \cap B) + c(A \cup B) \leq c(A) + c(B)$ ,  $\forall A, B \subseteq \Omega$ .

When a comonotonic risk measure is additionally law-invariant (i.e., it is a *distortion* measure), the Choquet capacities are uniquely determining by a concave distortion function, i.e.,

$$c(S) = \Psi(\mathbb{P}(S)), \forall S \in \mathcal{F}, \quad (2)$$

<sup>3</sup> It is known that comonotonicity actually implies positive homogeneity [25], so the we can define comonotonic risk measures as those satisfying [P1-3] and [P5] (the reverse is not true, i.e., not all coherent risk measures are comonotonic [3]).

where  $\Psi : [0, 1] \rightarrow [0, 1]$  is a concave, nondecreasing function satisfying  $\Psi(0) = 0$  and  $\Psi(1) = 1$ .

A popular example of comonotonic (in fact, distortion) risk measure, studied extensively in the literature, is Average Value-at-Risk at level  $\varepsilon \in [0, 1]$  ( $\text{AVaR}_\varepsilon$ ), also known as *Conditional Value-at-Risk*, *Tail Value-at-Risk* or *Expected Shortfall*. It is defined as

$$\text{AVaR}_\varepsilon(X) \stackrel{\text{def}}{=} \frac{1}{\varepsilon} \int_{1-\varepsilon}^1 \text{VaR}_{1-t}(X) dt. \quad (3a)$$

where  $\text{VaR}_\varepsilon(X) \stackrel{\text{def}}{=} \inf\{m \in \mathbb{R} : \mathbb{P}[X - m > 0] \leq \varepsilon\}$  is the Value at Risk at level  $\varepsilon$ . As the name suggests,  $\text{AVaR}_\varepsilon$  represents an average of all VaR measures with level at most  $\varepsilon$ . When the underlying reference measure  $\mathbb{P}$  is non-atomic, it can be shown [25] that  $\text{AVaR}_\varepsilon(X) = \mathbb{E}_\mathbb{P}[X | X \geq \text{VaR}_\varepsilon(X)]$ , which motivates the second and third names that the latter measure bears. While AVaR is a distortion measure, VaR is not even convex, since it fails requirement [P3].

**2.3. Dynamic Risk Measures.** As stated in the introduction, the main focus of the present paper are dynamic risk measures, i.e., risk measures defined for cash streams that are received or dispensed across several time-periods. More precisely, a dynamic risk measure entails the specification of an entire sequence of risk measures  $\{\mu_{[t,T]}\}_{t=0}^{T-1}$ , such that  $\mu_{[t,T]}$  maps a future stream of random costs  $X_{[t,T]} \stackrel{\text{def}}{=} (X_t, \dots, X_T)$  into risk assessments at  $t$ .

Following a large body of literature [40, 6, 21, 45, 15, 23, 48, 47, 16], we furthermore restrict the risk measurements at time  $t$  to only depend on the cumulative costs in the future, i.e., we take  $\mu_{[t,T]} : \mathcal{X}_T \rightarrow \mathcal{X}_t$ , and the risk of  $X_{[t,T]}$  is  $\mu_{[t,T]}(X_t + \dots + X_T)$ . While such measures have been criticized for ignoring the timing when future cashflows are received, they are consistent with the assumptions in many academic papers focusing on portfolio management under risk [10, 20], as well as with current risk management practice [35], and provide a natural, simpler first step in our analysis.

In this framework, we introduce the first way of measuring dynamic risk, whereby  $\mu_{[t,T]}$  is obtained by applying a static risk measure, conditioned on information available at time  $t$ . In the context of the probabilistic space of Section 2.1, this can be formalized as follows.

**DEFINITION 2.** A *time inconsistent (dynamic) risk measure* is any set of mappings  $\{\mu_{[t,T]}\}_{t=1}^T$  of the form  $\mu_{[t,T]} = (\mu^i)_{i \in \Omega_t}, \forall t \in [0, T]$ , where  $\mu^i : \mathcal{X}_T \rightarrow \mathbb{R}$  is a risk measure, for any node  $i \in \Omega_t$ .

In other words, conditional on reaching node  $i \in \Omega_t$  at time  $t$ , the risk of a future cashflow  $X_{[t,T]}$  is given by  $\mu^i(\sum_{\tau=t}^T X_\tau)$ , where every  $\mu^i$  is a static risk measure, which can be furthermore required to satisfy additional axiomatic properties, as per Section 2.2.

The choice above is eminently sensible - the specification of risk can be done in a unified fashion, by means of a single risk measure at every node and time. This makes for a compact representation of risk preferences, which can be more easily calibrated from empirical data, more readily comprehended and adopted by practitioners, and more uniformly applied across a variety of businesses and products. For instance, it is by far the most common paradigm in financial risk management, where a 10-day VaR is typically calculated at level  $\varepsilon = 1\%$ , assuming the trading portfolio remains fixed during the assessment period [35].

However, as the name suggests, such risk measures readily result in time inconsistent behavior. To see this, consider the following example, adapted from Roorda et al. [45].

**EXAMPLE 1.** Consider the tree in Figure 1, with the elementary events  $\Omega = \{UU, UD, DU, DD\}$ . Consider the risk measure given by

$$\mu^i(X) = \max_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_\mathbb{P}[X|i], \quad \forall i \in \{R, U, D\},$$

where  $\mathcal{P}$  contains two probability measures, one corresponding to  $p = 0.4$ , and the other to  $p = 0.6$ . Clearly, all  $\{\mu^i\}_{i \in \{R, U, D\}}$  correspond to coherent risk measures. For the random cost  $X$  such that

$X(UU) = X(DD) = 0$ , and  $X(UD) = X(DU) = 100$ , we have  $\mu^U(X) = \mu^D(X) = 60$ , and  $\mu^R(X) = 48$ . Therefore,  $X$  is deemed strictly riskier than a deterministic cost  $Y = 50$  in all states of nature at time  $t = 1$ , but nonetheless  $Y$  is deemed riskier than  $X$  at time  $t = 0$ .

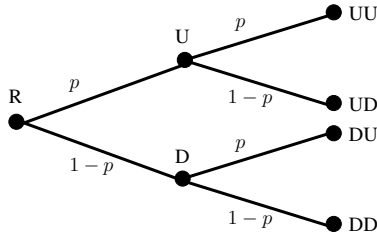


FIGURE 1. Example showing time inconsistency of a static risk measure. The random cost  $X$  with  $X(UU) = X(DD) = 0$ ,  $X(UD) = X(DU) = 100$  is deemed strictly riskier in all states of nature at time  $t = 1$  than a deterministic cost  $Y = 50$ , but nonetheless  $Y$  is deemed riskier than  $X$  at time  $t = 0$ .

We note that there is nothing peculiar in the choices above, in that similar counterexamples can be constructed for any risk measures  $\mu^i$ , even when the latter are comonotonic. Rather, the issue at play is the key feature distinguishing dynamic from static risk assessment, namely the consistency in the risk preference profile over time. This is summarized in the axiom<sup>4</sup> of *time* (or *dynamic*) *consistency*, which asks that a dynamic risk measure  $\{\mu_{t,T}\}_{t=0}^{T-1}$  should satisfy, for all  $t \in [0, T-1]$  and all  $X, Y \in \mathcal{X}_T$ ,

$$\mu_{t+1,T}(X) \geq \mu_{t+1,T}(Y) \text{ implies } \mu_{t,T}(X) \geq \mu_{t,T}(Y).$$

This is a requirement on the particular functional forms that can be considered for  $\mu_{[t,T]}$ , which is typically violated by the naïve dynamic measures of Definition 2. A central result in the literature [40, 6, 21, 45, 15, 46, 39, 23, 47] is the following theorem, stating that any consistent measure has a compositional representation in terms of one-period risk mappings.

**THEOREM 2.** *Any dynamic risk measure  $\{\mu_{t,T}\}_{t=0}^{T-1}$  that is time consistent can be written as*

$$\mu_{t,T}(X_t + \dots + X_T) = \mu_{t+1} \left( \mu_{t+2} \left( \dots \left( \mu_T(X_t + \dots + X_T) \right) \dots \right) \right). \quad (4)$$

where  $\mu_t : \mathcal{X}_t \rightarrow \mathcal{X}_{t-1}$ ,  $t \in [1, T]$  are a set of single-period conditional risk mappings.

This leads us to define the second way of measuring dynamic risk on the scenario tree of Section 2.1, by means of composing risk measures.

**DEFINITION 3.** A set of mappings  $\{\mu_{[t,T]}\}_{t=0}^{T-1}$  is said to be a *time consistent (dynamic) risk measure* if  $\mu_{[t,T]} = \mu_{t+1} \circ \mu_{t+2} \circ \dots \circ \mu_T$  for any  $t \in [0, T-1]$ , where  $\mu_{t+1} \equiv (\mu^i)_{i \in \Omega_t}$ , and  $\mu^i : \mathbb{R}^{|\mathcal{C}_i|} \rightarrow \mathbb{R}$  are risk measures, for any  $i \in \Omega_t$ .

We say that  $\{\mu_{[t,T]}\}_{t=0}^{T-1}$  is a time-consistent, coherent (comonotonic) risk measure if every  $\mu^i$  is coherent (respectively, comonotonic), for any  $i \in \Omega_t$  and  $t \in [0, T-1]$ .

Apart from being axiomatically justified, this compositional form has the advantage of allowing a recursive estimation of the risk, and an application of the Bellman optimality principle in optimization problems involving dynamic risk measures [38, 34, 48, 47]. This has led to its adoption in

<sup>4</sup>We note that there are several notions of time consistency in the literature (see Penner [39], Acciaio and Penner [1], Roorda and Schumacher [46] for an in-depth discussion and comparison). The one we adopt here is closest in spirit to strong dynamic consistency, and seems to be the most widely accepted notion in the literature.

actuarial science [30, 14], as well as in several recent papers that re-examine operational problems under coherent measures of risk [4, 17].

The main downside of the compositional form is that it requires a specification of all the mappings  $\mu^i$ , which furthermore no longer lends itself to an easy interpretation, particularly as seen from the perspective of time  $t = 0$ . In particular, even if  $\mu^i$  corresponded to the same primitive risk measure  $\mu$ , the overall compositional measure<sup>5</sup>  $\mu_{[0,T]} = \mu \circ \mu \circ \dots \circ \mu$  would bear no immediate relation to  $\mu$ . As an example, when  $\mu^i = \text{AVaR}_\varepsilon$ , the overall  $\mu_{[0,T]}$  corresponds to the so-called ‘‘iterated CTE’’ [30, 14, 46], which does not lend itself to the same simple interpretation as a single AVaR. Furthermore, practitioners often feel that the overall risk measure  $\mu_{[0,T]}$  is *overly conservative*, since it composes what are already potentially conservative risk evaluations backwards in time – for instance, for the iterated TCE, one is taking tail conditional expectations of tail conditional expectations. This has been recognized informally in the literature by Roorda and Schumacher [46, 44], who propose new notions of time consistency that avoid the issue, but without establishing precisely whether or to what extent the conservatism is actually true. From a different perspective, it is not obvious how ‘‘close’’ a particular compositional measure is to a given inconsistent one, and how one could go about constructing the latter in a way that tightly approximates the former.

**2.4. Main Problem Statement.** In this context, the goal of the present paper is to take the first step towards better understanding the tradeoffs between the two ways of measuring risk. More precisely, we consider dynamic risk as viewed from the perspective of time  $t = 0$ , and examine two potential metrics: a time-inconsistent (comonotonic) risk measure  $\mu_I : \mathcal{X}_T \rightarrow \mathbb{R}$ , and a time-consistent (comonotonic) risk measure  $\mu_C : \mathcal{X}_T \rightarrow \mathbb{R}$ . For two such metrics, we seek to address the following related problems.

PROBLEM 1. Given  $\mu_I$  and  $\mu_C$ , test whether

$$\mu_C(Y) \leq \mu_I(Y), \forall Y \in \mathcal{X}_T \quad \text{or} \quad \mu_I(Y) \leq \mu_C(Y), \forall Y \in \mathcal{X}_T.$$

PROBLEM 2. Given  $\mu_I, \mu_C$ , find the smallest  $\alpha_{\mu_C, \mu_I} > 0$  and  $\alpha_{\mu_I, \mu_C} > 0$  such that

$$\text{if } \mu_C(Y) \leq \mu_I(Y), \forall Y \quad \mu_I(Y) \leq \alpha_{\mu_C, \mu_I} \cdot \mu_C(Y), \forall Y \in \mathcal{X}_T, Y \geq 0 \quad (5a)$$

$$\text{if } \mu_I(Y) \leq \mu_C(Y), \forall Y \quad \mu_C(Y) \leq \alpha_{\mu_I, \mu_C} \cdot \mu_I(Y), \forall Y \in \mathcal{X}_T, Y \geq 0. \quad (5b)$$

A satisfactory answer to Problem 1 would provide a test for whether one of the formulations is always over or under estimating risk as compared to the other. As we show, consistent measures obtained by iterating the same primitive measure  $\mu$  do not necessarily over (or under) estimate risk as compared to  $\mu$ , and this is true even when  $\mu = \text{AVaR}_\varepsilon$ , the case considered in [46, 44]. However, by composing  $\mu$  with conditional expectation operators, one always obtains lower bounds to the static risk measurement under  $\mu$ . For instance,  $\mu \circ \mathbb{E}$  and  $\mathbb{E} \circ \mu$  are both lower bounds to a given static evaluation by  $\mu$ . By contrast, upper bounds are obtained only when composing with worst-case operators in the *final* periods of the horizon: e.g.,  $\mu \circ \max$  is necessarily an upper bound for  $\mu$ , but  $\max \circ \mu$  is not.

To understand the relevance of Problem 2, note that the minimal factors  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  satisfying (5a) and (5b), respectively, provide a compact notion of how closely  $\mu_I$  can be approximated through lower or upper bounding consistent measures  $\mu_C$ , respectively. Since, in practice, it may be far easier to elicit or estimate a single static risk measure  $\mu_I$ , characterizing and computing  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  constitutes the first step towards constructing the time-consistent risk measure  $\mu_C$  that is ‘‘closest’’ to a given  $\mu_I$ . We note that a similar concept of inner and outer approximations by

<sup>5</sup> Here and throughout the paper, we use the shorthand notation  $\mu \circ \mu$  with the understanding that the elementary risk measure  $\mu$  is applied in stages  $t \geq 1$  in a conditional fashion.



means of distortion risk measures appears in [12]. However, the goal and analysis there are quite different, since the question is to approximate a static risk measure by means of another static distortion risk measure.

In a different sense, the smallest  $\alpha_{\mu_I, \mu_C}^*$  could be used to scale up risk measurements according to  $\mu_I$  in order to turn them into “safe” (i.e., conservative) estimates of measurements according to  $\mu_C$ . Scaling risk assessments by particular factors is actually common practice in financial risk management: according to the rules set forth by the Basel Committee for banking regulation, banks are required to report the 10-day VaR calculated at 1% level, which is then *multiplied by a factor of 3* to provide the minimum capital requirement for regulatory purposes; the factor of 3 is meant to account for losses occurring beyond VaR, and also for potential model misspecification (see Chapter 5 of [35]). Therefore, this usage of  $\alpha_{\mu_I, \mu_C}^*$  could integrate well with practice.

We conclude the section by a remark pertinent to the two problems and our analysis henceforth.

**REMARK 1.** On first sight, the requirement of non-negative  $Y$  in the text of Problem 2 might seem overly restrictive. However, note that, if we insisted on  $\mu_C(Y) \leq \mu_I(Y)$  holding for *any* cost  $Y$ , and if  $\mu_C, \mu_I$  were allowed to take both positive and negative values, then the questions in Problem 2 would be meaningless, in that no feasible  $\alpha_{\cdot, \cdot}$  values would exist satisfying (5a) or (5b). To this end, we are occasionally forced to make the assumption that the stochastic losses  $Y$  are non-negative. This is not too restrictive whenever a lower bound  $Y_L$  is available for  $Y$ . By using the cash-invariance property ([P2]) of the risk measures involved, one could reformulate the original question with regards to the random loss  $Y - Y_L$ , which would be nonnegative. Furthermore, in specific applications (e.g., multi-period inventory management [4]),  $Y$  is the sum of intra-period costs  $X_t$  that are always non-negative, so requiring  $Y$  to be nonnegative is quite sensible.

**3. Bounds for Coherent and Comonotonic Risk Measures.** In this section, we seek appropriate answers to Problem 1 and Problem 2, with the end-goal of characterizing the tightest multiplicative approximation of a given inconsistent risk measure by means of lower (or upper) bounding consistent risk measures.

To keep the discussion compact and avoid repetitive arguments, we first treat an abstract setting of comparing two coherent measures on the same space of outcomes. The conditions that we derive will be quite general, since no further structure will be imposed on the two measures. Section 3.2 will then discuss in detail the comparison between a *time-inconsistent, comonotonic* risk measure  $\mu_I$  and a *consistent, comonotonic* risk measure  $\mu_C$ , deriving particular forms for the results and conditions. In Section 3.3, we derive the main technical result needed to obtain multiplicative bounds on inconsistent, comonotonic risk measures, which we then use in Section 3.2 to address the main problems of interest.

Consider a discrete probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , and let  $\mathcal{X}$  be the space of all random variables on  $\Omega$  (isomorphic with  $\mathbb{R}^{|\Omega|}$ ). On this space, we are interested in comparing two *coherent* risk measures  $\mu_{1,2} : \mathcal{X} \rightarrow \mathbb{R}$  given by *polyhedral* sets of measures, i.e.,

$$\mu_i(Y) = \max_{\mathbf{q} \in \mathcal{Q}_i} \mathbf{q}^T \mathbf{Y}, \quad \forall Y \in \mathcal{X}, \quad \forall i \in \{1, 2\},$$

where  $\mathcal{Q}_{1,2} \subseteq \Delta^{|\Omega|}$  are (bounded) polyhedra<sup>6</sup>. Our main focus is on (1) characterizing conditions such that  $\mu_1(Y) \leq \mu_2(Y), \forall Y \in \mathcal{X}$ , and (2) finding the smallest factor  $\alpha$  such that

$$\mu_1(Y) \leq \mu_2(Y) \leq \alpha \mu_1(Y), \quad \forall Y \in \mathcal{X} \quad (Y \geq 0).$$

<sup>6</sup> Several of the results discussed here readily extend to arbitrary closed, convex sets of representing measures. We restrict attention to the polyhedral case since it captures the entire class of comonotonic risk measures, it is simpler to describe, and computationally advantageous, since evaluating the risk measure entails solving a linear program.

In this context, the risk measure  $\mu_i(Y)$  can be identified as the support function of the convex set  $\mathcal{Q}_i$ , so that the following standard result in convex analysis (see, e.g., [42, Corollary 13.1.1]) can be invoked to test whether one risk measure dominates the other.

PROPOSITION 1. *The inequality  $\mu_1(Y) \leq \mu_2(Y)$ ,  $\forall Y \in \mathcal{X}$  holds if and only if  $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$ .*

The usefulness of the latter condition critically depends on the representation of the sets of measures  $\mathcal{Q}_i$ . For instance, if  $\mathcal{Q}_i$  are polytopes, the containment problem  $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$  is co-NP-complete when  $\mathcal{Q}_1$  is given by linear inequalities and  $\mathcal{Q}_2$  is given by its extreme points, but it can be solved in polynomial time, by linear programming (LP), for all the other three possible cases [26].

Proposition 1 also sheds light on the second question of interest, through the following corollary.

COROLLARY 1. *There does not exist any  $\alpha \neq 1$  such that  $\mu_2(Y) \leq \alpha \mu_1(Y)$ ,  $\forall Y \in \mathcal{X}$ .*

*Proof.* For any  $\alpha > 0$ , the condition  $\{\mu_2(Y) \leq \alpha \mu_1(Y), \forall Y \in \mathcal{X}\}$  is equivalent to  $\mathcal{Q}_2 \subseteq \alpha \mathcal{Q}_1$ . Since  $\mathcal{Q}_{1,2} \subseteq \Delta^{|\Omega|}$ , the containment cannot hold for any  $\alpha \neq 1$ .  $\square$

This result prompts the need to restrict the space of random losses considered. As suggested in Section 2.4, an eminently sensible choice is to take  $Y \geq 0$ , which is always reasonable when a lower bound on the losses is available. This allows us to characterize the desired conditions by examining inclusions of *down-monotone closures* of the sets  $\mathcal{Q}_i$ . To this end, we introduce the following two definitions (see Section A of the Appendix for more details and references).

DEFINITION 4. A non-empty set  $Q \subseteq \mathbb{R}_+^n$  is said to be *down-monotone* if for any  $\mathbf{x} \in Q$  and any  $\mathbf{y}$  such that  $0 \leq \mathbf{y} \leq \mathbf{x}$ , we also have  $\mathbf{y} \in Q$ .

DEFINITION 5. The *down-monotone closure* of a set  $Q \subseteq \mathbb{R}_+^n$ , denoted by  $\text{sub}(Q)$ , is the smallest down-monotone set containing  $Q$ , i.e.,

$$\text{sub}(Q) \stackrel{\text{def}}{=} \{ \mathbf{x} \in \mathbb{R}_+^n : \exists \mathbf{q} \in Q, \mathbf{x} \leq \mathbf{q} \}.$$

When restricting attention to nonnegative losses, one can readily show that a coherent risk measure can be obtained by evaluating the worst case over an *extended* set of generalized scenarios, given by the down-monotone closure of the original set. This is summarized in the following extension of representation Theorem 1.

PROPOSITION 2. *Let  $\mu(Y) = \max_{\mathbf{q} \in \mathcal{Q}} \mathbf{q}^T \mathbf{Y}$  be a coherent risk measure. Then,*

$$\mu(Y) = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q})} \mathbf{q}^T \mathbf{Y}, \forall Y \geq 0. \quad (6)$$

*Proof.* The inequality  $\max_{\mathbf{q} \in \mathcal{Q}} \mathbf{q}^T \mathbf{Y} \leq \max_{\mathbf{q} \in \text{sub}(\mathcal{Q})} \mathbf{q}^T \mathbf{Y}$  follows simply because  $\mathcal{Q} \subseteq \text{sub}(\mathcal{Q})$ . To prove the reverse, consider any  $Y \geq 0$  and let  $\mathbf{q}_1$  be the maximizer of  $\max_{\mathbf{q} \in \text{sub}(\mathcal{Q})} \mathbf{q}^T \mathbf{Y}$ . By Definition 5, there exists  $\mathbf{q}_2 \in \mathcal{Q}$  such that  $\mathbf{q}_2 \geq \mathbf{q}_1 \geq 0$ . Then:

$$\max_{\mathbf{q} \in \mathcal{Q}} \mathbf{q}^T \mathbf{Y} \geq \mathbf{q}_2^T \mathbf{Y} \geq \mathbf{q}_1^T \mathbf{Y} = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q})} \mathbf{q}^T \mathbf{Y}. \quad \square$$

In view of this result, one can readily show that testing whether a risk measurement dominates another can be done *equivalently* in terms of the down-monotone closures of the representing sets of measures, as stated in the next result.

LEMMA 1. *The inequality  $\mu_1(Y) \leq \mu_2(Y)$ ,  $\forall Y \in \mathcal{X}$  holds if and only if  $\text{sub}(\mathcal{Q}_1) \subseteq \text{sub}(\mathcal{Q}_2)$ .*

*Proof.* By Proposition 1, the above is equivalent to showing

$$\mathcal{Q}_1 \subseteq \mathcal{Q}_2 \Leftrightarrow \text{sub}(\mathcal{Q}_1) \subseteq \text{sub}(\mathcal{Q}_2).$$

( $\Rightarrow$ ) Consider  $\mathbf{q}_1 \in \text{sub}(\mathcal{Q}_1)$ . Then, by Definition 5, there exists  $\mathbf{q}'_1 \in \mathcal{Q}_1$  such that  $\mathbf{q}'_1 \geq \mathbf{q}_1$ . Since  $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$ , we have  $\mathbf{q}'_1 \in \mathcal{Q}_2$ , and therefore  $\mathbf{q}_1 \in \text{sub}(\mathcal{Q}_2)$ .

( $\Leftarrow$ ) Note that  $\mathcal{Q}_i = \text{sub}(\mathcal{Q}_i) \cap \Delta^{|\Omega|}$  for  $i = 1, 2$ . Then:

$$\begin{aligned} \text{sub}(\mathcal{Q}_1) \subseteq \text{sub}(\mathcal{Q}_2) &\Rightarrow \\ \text{sub}(\mathcal{Q}_1) \cap \Delta^{|\Omega|} \subseteq \text{sub}(\mathcal{Q}_2) \cap \Delta^{|\Omega|} &\Leftrightarrow \\ \mathcal{Q}_1 \subseteq \mathcal{Q}_2. &\quad \square \end{aligned}$$

Figure 2 depicts examples of the sets  $\mathcal{Q}_1, \mathcal{Q}_2$  and their down-monotone closures  $\text{sub}(\mathcal{Q}_1)$  and  $\text{sub}(\mathcal{Q}_2)$ , respectively. Note that the conditions provided by Lemma 1 hold for *any* cost  $Y$ , i.e., non-negativity is not needed; they may also be more efficient in practice than directly checking  $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$ , in cases when a suitable representation is available for  $\text{sub}(\mathcal{Q}_{1,2})$ , but not for  $\mathcal{Q}_{1,2}$  [26].

By considering down-monotone closures and restricting to nonnegative losses, we can also address the second question of interest, namely retrieving the smallest scaling factor  $\alpha$  such that  $\mu_2(Y) \leq \alpha \cdot \mu_1(Y)$ . The following result characterizes any such feasible  $\alpha$ .

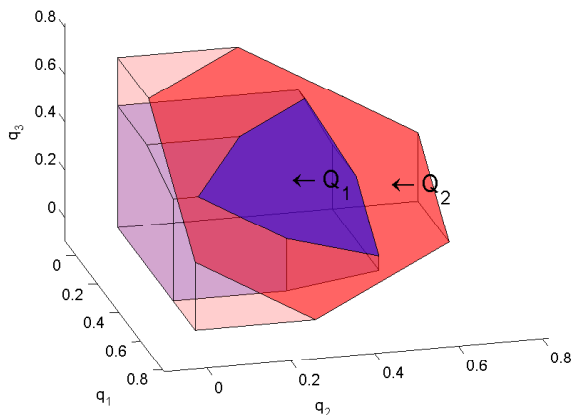


FIGURE 2. Inclusion relation between  $\mathcal{Q}_1, \mathcal{Q}_2$  (and the corresponding down-monotone closures,  $\text{sub}(\mathcal{Q}_1)$  and  $\text{sub}(\mathcal{Q}_2)$ , respectively) that is equivalent to  $\mu_1(Y) \leq \mu_2(Y)$ ,  $\forall Y \in \mathcal{X}_2$ .

PROPOSITION 3. *The inequality  $\mu_2(Y) \leq \alpha \cdot \mu_1(Y)$  holds for all  $Y \geq 0$  if and only if  $\text{sub}(\mathcal{Q}_2) \subseteq \alpha \cdot \text{sub}(\mathcal{Q}_1)$ .*

*Proof.* By Proposition 2,  $\mu_i(Y) = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_i)} \mathbf{q}^T Y$  for  $i = 1, 2$ . The claim then follows directly from Lemma 1 in [29] (also see Theorem 11 in the Appendix).  $\square$

In view of this result, the *minimal*  $\alpha$  exactly corresponds to the smallest inflation of the down-monotone polytope  $\text{sub}(\mathcal{Q}_1)$  that contains the down-monotone polytope  $\text{sub}(\mathcal{Q}_2)$ . This identification leads to the following characterization of the optimal scaling factor.

THEOREM 3. *Let  $\alpha_{\mu_1, \mu_2}^*$  denote the smallest value of  $\alpha$  such that  $\mu_2(Y) \leq \alpha \cdot \mu_1(Y)$ ,  $\forall Y \geq 0$ .*

1. *If  $\text{sub}(\mathcal{Q}_1) = \{\mathbf{q} \in \mathbb{R}_+^n : \mathbf{a}_i^T \mathbf{q} \leq b_i, \forall i \in \mathcal{I}\}$ , where  $\mathbf{a}_i \geq \mathbf{0}$ ,  $b_i \geq 0$ , then*

$$\alpha_{\mu_1, \mu_2}^* = \max_{i \in \mathcal{I}} \frac{\max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_2)} \mathbf{a}_i^T \mathbf{q}}{b_i} = \max_{i \in \mathcal{I}} \frac{\max_{\mathbf{q} \in \mathcal{Q}_2} \mathbf{a}_i^T \mathbf{q}}{b_i} = \max_{i \in \mathcal{I}} \frac{\mu_2(\mathbf{a}_i)}{b_i}. \quad (7)$$

2. If  $\mathcal{Q}_1 = \{\mathbf{q} \in \mathbb{R}^n : A\mathbf{q} \leq \mathbf{b}\}$ , then  $\alpha_{\mu_1, \mu_2}^*$  is the smallest value  $t$  such that the optimal value of the following bilinear program is at most zero:

$$\begin{aligned} \max_{\mathbf{q}, \boldsymbol{\mu}} \quad & (A\mathbf{q} - t\mathbf{b})^T \boldsymbol{\mu} \\ & \mathbf{q} \in \mathcal{Q}_2 \\ & \boldsymbol{\mu} \geq \mathbf{0} \\ & A^T \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{8}$$

*Proof.* The first claim is a known result in combinatorial optimization – see Theorem 11 in the Appendix and Theorem 2 in [29] for a complete proof.

To argue the second claim, note that the smallest  $\alpha$  can be obtained, by definition, as follows:

$$\begin{aligned} \min \left\{ t : \max_{\mathbf{Y} \geq \mathbf{0}} \left[ \max_{\mathbf{q} \in \mathcal{Q}_2} \mathbf{Y}^T \mathbf{q} - t \cdot \max_{\mathbf{q} \in \mathcal{Q}_1} \mathbf{Y}^T \mathbf{q} \right] \leq 0 \right\} &= \min \left\{ t : \max_{\mathbf{Y} \geq \mathbf{0}} \left[ \max_{\mathbf{q} \in \mathcal{Q}_2} \mathbf{Y}^T \mathbf{q} - t \cdot \min_{\substack{A^T \boldsymbol{\mu} = \mathbf{Y} \\ \boldsymbol{\mu} \geq \mathbf{0}}} \mathbf{b}^T \boldsymbol{\mu} \right] \leq 0 \right\} \\ &= \min \left\{ t : \max_{\substack{\mathbf{q} \in \mathcal{Q}_2 \\ \boldsymbol{\mu} \geq \mathbf{0} \\ A^T \boldsymbol{\mu} \geq \mathbf{0}}} (A\mathbf{q} - t\mathbf{b})^T \boldsymbol{\mu} \leq 0 \right\}. \end{aligned}$$

The first equality follows by strong LP duality applied to the maximization over  $\mathbf{q} \in \mathcal{Q}_1$ , which always has a finite optimum since  $\mathcal{Q}_1$  is bounded. The second equality follows by replacing the inner minimization with a maximization, switching the order of the maximizations, and eliminating the variables  $\mathbf{Y}$ .  $\square$

The results in Theorem 3 give a direct connection between the problem of computing  $\alpha_{\mu_1, \mu_2}^*$  and the representations available for the sets  $\mathcal{Q}_i$  and  $\text{sub}(\mathcal{Q}_i)$ . More precisely,

- If a polynomially-sized inequality description is available for  $\text{sub}(\mathcal{Q}_1)$ , then  $\alpha_{\mu_1, \mu_2}^*$  can be obtained by solving the small number of LPs in (7). Every such LP essentially entails an evaluation of the risk measure  $\mu_2$ , leading to an efficient overall procedure.
- If a compact inequality representation is available for  $\mathcal{Q}_1$ , then  $\alpha_{\mu_1, \mu_2}^*$  can be found by bisection search over  $t \geq 0$ , where in each step the bilinear program in (8) is solved. Since bilinear programs can be reformulated as integer programs [31], for which powerful commercial solvers are available, this approach may lead to a scalable procedure, albeit not one with polynomial-time complexity.

Our observations concerning the complexity of testing  $\text{sub}(\mathcal{Q}_2) \subseteq \alpha \cdot \text{sub}(\mathcal{Q}_1)$  are summarized in Table 1 below. When  $\mathcal{Q}_2$  or  $\text{sub}(\mathcal{Q}_2)$  have polynomially-sized vertex descriptions, the test simply requires checking containment for a finite set of points, and when  $\text{sub}(\mathcal{Q}_2)$  has a polynomially-sized description, the results of Theorem 3 apply. We conjecture that all the remaining cases are NP-complete, but do not pursue a formal analysis in the present paper. Section 4.1 revisits the question of computational complexity in the context of *comonotonic* risk measures, and argues that the general containment problem is NP-hard.

	Poly ext( $\mathcal{Q}_1$ )	Poly face( $\mathcal{Q}_1$ )	Poly ext(sub( $\mathcal{Q}_1$ ))	Poly face(sub( $\mathcal{Q}_1$ ))
Poly ext( $\mathcal{Q}_2$ )	P	P	P	P
Poly face( $\mathcal{Q}_2$ )				P
Poly ext(sub( $\mathcal{Q}_2$ ))	P	P	P	P
Poly face(sub( $\mathcal{Q}_2$ ))				P

TABLE 1. Computational complexity for determining whether  $\mu_2(Y) \leq \alpha \cdot \mu_1(Y)$ ,  $\forall Y \geq \mathbf{0}$  for a given  $\alpha > 0$ . “Poly ext” and “Poly face” denote a polynomially-sized vertex and inequality description, respectively. “P” denotes a polynomial-time algorithm is available.

We conclude our general discussion by noting that the tightest scaling factor  $\alpha_{\mu_1, \mu_2}^*$  can also be used to directly re-examine the first question of interest, namely testing when a given coherent

risk measure upper bounds another. This is formalized in the following corollary, which is a direct result of Lemma 1 and Proposition 3.

**COROLLARY 2.** *The inequality  $\mu_2(Y) \leq \mu_1(Y)$ ,  $\forall Y \in \mathcal{X}$  holds if and only if  $\alpha_{\mu_1, \mu_2}^* \leq 1$ .*

The latter result suggests that characterizing and computing the tightest scaling factor is instrumental in answering *all* questions relating to the approximation of a coherent risk measure by means of another. In particular, given  $\mu_1$  and  $\mu_2$ , by determining the scaling factors  $\alpha_{\mu_1, \mu_2}^*$  and  $\alpha_{\mu_2, \mu_1}^*$ , we can readily test domination and also approximate one measure by the other, as follows:

- If  $\alpha_{\mu_2, \mu_1}^* \leq 1$ , then  $\mu_1(Y) \in \left[ \frac{1}{\alpha_{\mu_1, \mu_2}^*}, 1 \right] \cdot \mu_2(Y)$ ,  $\forall Y \geq 0$ .
- If  $\alpha_{\mu_1, \mu_2}^* \leq 1$ , then  $\mu_1(Y) \in \left[ 1, \alpha_{\mu_2, \mu_1}^* \right] \cdot \mu_2(Y)$ ,  $\forall Y \geq 0$ .

**3.1. Tightest Time Consistent and Coherent Upper Bound.** The results and exposition in the prior section made no reference to the way in which the coherent risk measures  $\mu_{1,2}$  were obtained, as long as the sets of representing measures  $\mathcal{Q}_1, \mathcal{Q}_2$  were polyhedral. In this section, we discuss some of these results in the context of Section 2.4 – more precisely, we take  $\mu_1$  as the time-inconsistent risk measure  $\mu_I$ , while  $\mu_2$  denotes the compositional measure  $\mu_C$ .

Our goal is to show that, when  $\mu_I$  is coherent, a complete characterization of the tightest possible uniform upper bound to  $\mu_I$  is readily available, and is given by a popular construction in the literature [22, 45, 6, 54]. This not only yields the tightest possible factor  $\alpha_{\mu_I, \mu_C}^*$ , but also considerably simplifies the test  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y$ , for any coherent  $\mu_C$ .

The next proposition introduces this construction for an arbitrary coherent measure  $\mu$ .

**PROPOSITION 4.** *Consider a risk measure  $\mu(Y) = \sup_{\mathbf{q} \in \mathcal{Q}} \mathbf{q}^T \mathbf{Y}$ ,  $\forall Y \in \mathcal{X}_T$ , and define the risk measure  $\hat{\mu}(Y) \stackrel{\text{def}}{=} (\hat{\mu}_1 \circ \hat{\mu}_2 \circ \dots \circ \hat{\mu}_T)(Y)$ , where the mappings  $\hat{\mu}_t \equiv (\hat{\mu}^i)_{i \in \Omega_{t-1}} : \mathcal{X}_t \rightarrow \mathcal{X}_{t-1}$  are given by*

$$\forall t \in [1, T], \forall i \in \Omega_{t-1}, \quad \hat{\mu}^i(Y) \stackrel{\text{def}}{=} \sup_{\mathbf{q} \in \hat{\mathcal{Q}}_\mu^i} \mathbf{q}^T \mathbf{Y}, \quad \forall \mathbf{Y} \in \mathbb{R}^{|\mathcal{C}_i|}, \quad (9)$$

$$\hat{\mathcal{Q}}_\mu^i \stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \Delta^{|\mathcal{C}_i|} : \exists \mathbf{p} \in \mathcal{Q} : q_j = \frac{\mathbf{p}(\mathcal{D}_j)}{\mathbf{p}(\mathcal{D}_i)}, \forall j \in \mathcal{C}_i \right\}. \quad (10)$$

*Then,  $\hat{\mu}$  is a time-consistent, coherent risk measure, and  $\mu(Y) \leq \hat{\mu}(Y)$ ,  $\forall Y \in \mathcal{X}_T$ .*

As mentioned, this construction has already been considered in several papers in the literature, and several authors have recognized that it provides an upper bound to  $\mu$ . It is known that  $\hat{\mu}$  is time-consistent, and has a representation of the form  $\hat{\mu}(Y) = \sup_{\mathbf{q} \in \hat{\mathcal{Q}}_\mu} \mathbf{q}^T \mathbf{Y}$ ,  $\forall Y \in \mathcal{X}_T$ , where the set  $\hat{\mathcal{Q}}_\mu$  has a *product* or *rectangular* structure. Note that it is obtained by computing products of the sets  $\hat{\mathcal{Q}}_\mu^i$  of single-step conditional probabilities obtained by marginalization at each node in the tree. Furthermore,  $\mathcal{Q} \subseteq \hat{\mathcal{Q}}_\mu$ , and therefore  $\mu(Y) \leq \hat{\mu}(Y)$ ,  $\forall Y \in \mathcal{X}_T$  [22, 45, 54].

**EXAMPLE 1 REVISITED.** *To understand the construction, consider again Example 1. The set  $\mathcal{Q}$  yielding the inconsistent measure  $\mu$  at the root node  $R$  is given by two probabilities, corresponding to  $p = 0.4$  and  $p = 0.6$ , i.e.,*

$$\mathcal{Q} = \{ (0.16, 0.24, 0.24, 0.36), (0.36, 0.24, 0.24, 0.16) \}.$$

*The sets of conditional one-step probabilities corresponding to nodes  $U, D$ , and  $R$  are then:*

$$\hat{\mathcal{Q}}_\mu^U = \hat{\mathcal{Q}}_\mu^D = \hat{\mathcal{Q}}_\mu^R = \{ (0.4, 0.6), (0.6, 0.4) \}.$$

*This yields a set  $\hat{\mathcal{Q}}_\mu$  containing eight different probability measures, for all possible products of one-step measures chosen from  $\hat{\mathcal{Q}}_\mu^U, \hat{\mathcal{Q}}_\mu^D, \hat{\mathcal{Q}}_\mu^R$ . More precisely,*

$$\hat{\mathcal{Q}}_\mu = \left\{ (0.16, 0.24, 0.24, 0.36), (0.16, 0.24, 0.36, 0.24), (0.24, 0.16, 0.24, 0.36), (0.24, 0.16, 0.36, 0.24), \right. \\ \left. (0.24, 0.36, 0.16, 0.24), (0.24, 0.36, 0.24, 0.16), (0.36, 0.24, 0.16, 0.24), (0.36, 0.24, 0.24, 0.16) \right\}.$$

In this context, we claim that  $\widehat{\mu}_I$  actually represents the *tightest* upper bound for  $\mu_I$ , among all possible coherent and time-consistent upper bounds. This is formalized in the following result.

LEMMA 2. *Consider any risk measure  $\mu_I(Y) = \sup_{\mathbf{q} \in \mathcal{Q}_I} \mathbf{q}^T \mathbf{Y}$ ,  $\forall Y \in \mathcal{X}_T$ , and let  $\widehat{\mu}_I$  be the corresponding risk measure obtained by the construction in Proposition 4. Also, consider any time-consistent, coherent risk measure  $\mu_C(Y) \stackrel{\text{def}}{=} (\rho_1 \circ \rho_2 \cdots \circ \rho_T)(Y)$ , where  $\rho_t \equiv (\rho_t^i)_{i \in \Omega_{t-1}} : \mathcal{X}_t \rightarrow \mathcal{X}_{t-1}$  are given by*

$$\rho_t^i(Y) = \max_{\mathbf{q} \in \mathcal{Q}_\rho^i} \mathbf{q}^T \mathbf{Y}, \forall \mathbf{Y} \in \mathbb{R}^{|\mathcal{C}_i|},$$

for some closed and convex sets  $\mathcal{Q}_\rho^i \subseteq \Delta^{|\mathcal{C}_i|}$ . Then, the following results hold:

1. If  $\mu_C(Y) \geq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_T$ , then

$$\mu_C(Y) \geq \widehat{\mu}_I(Y), \forall Y \in \mathcal{X}_T \quad \text{and} \quad \alpha_{\mu_I, \widehat{\mu}}^* \leq \alpha_{\mu_I, \mu_C}^*. \quad (11)$$

2.  $\mu_C(Y) \geq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_T$  holds if and only if

$$\widehat{\mathcal{Q}}_{\mu_I}^i \subseteq \mathcal{Q}_\rho^i, \forall i \in \Omega_{t-1}, \forall t \in [1, T]. \quad (12)$$

*Proof.* [1] Since  $\mu_C$  is a coherent risk measure, it can always be written as  $\mu_C(Y) = \max_{\mathbf{q} \in \mathcal{Q}_C} \mathbf{q}^T \mathbf{Y}$ . Furthermore, it is known that the set of representing measures  $\mathcal{Q}_C$  is obtained by taking products of the sets  $\mathcal{Q}_\rho^i$  (see, e.g., Roorda et al. [45] or Föllmer and Schied [25]). Due to this property,  $\mathcal{Q}_C$  is closed under the operation of taking marginals and computing the product of the resulting sets of conditional one-step measures [22, 45, 6], i.e.,

$$\mathcal{Q}_\rho^i = \left\{ \mathbf{q} \in \Delta^{|\mathcal{C}_i|} : \exists \mathbf{p} \in \mathcal{Q}_C : q_j = \frac{\mathbf{p}(\mathcal{D}_j)}{\mathbf{p}(\mathcal{D}_i)}, \forall j \in \mathcal{C}_i \right\}. \quad (13)$$

Since  $\mu_C(Y) \geq \mu_I(Y)$ , we must have  $\mathcal{Q}_I \subseteq \mathcal{Q}_C$ . But then, from (10) and (13), we obtain that  $\widehat{\mathcal{Q}}_{\mu_I}^i \subseteq \mathcal{Q}_\rho^i$ ,  $\forall i \in \Omega_{t-1}$ ,  $\forall t$ . This readily implies that  $\widehat{\mathcal{Q}}_{\mu_I} \subseteq \mathcal{Q}_C$ , and hence  $\widehat{\mu}_I(Y) \leq \mu_C(Y)$ ,  $\forall Y$ . The inequality for the multiplicative factors  $\alpha_{\mu_I, \cdot}^*$  follows from the definition.

[2] For the second result, note that the “ $\Rightarrow$ ” implication has already been proved in the first part. The reverse direction follows trivially since  $\widehat{\mathcal{Q}}_{\mu_I}^i \subseteq \mathcal{Q}_\rho^i$  implies that  $\widehat{\mathcal{Q}}_{\mu_I} \subseteq \mathcal{Q}_C$ , and, since  $\mu_I(Y) \leq \widehat{\mu}_I(Y)$ , we have  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y$ .  $\square$

The result above is useful in several ways. First, it suggests that the tightest time-consistent, coherent upper bound for a given  $\mu_I$  is  $\widehat{\mu}_I$ . This not only yields the smallest possible multiplicative factor  $\alpha_{\mu_I, \cdot}^*$ , but the upper-bound is *uniform*, i.e., for any loss  $Y$ . Also,  $\alpha_{\mu_I, \widehat{\mu}_I}^*$  is a lower bound on the best possible  $\alpha_{\mu_I, \mu_C}^*$  when the consistent measures  $\mu_C$  are further constrained, e.g., to be *comonotonic*.

The conditions (13) also prescribe a different way of testing  $\mu_I \leq \mu_C$ , by examining several smaller-dimensional tests involving the sets  $\widehat{\mathcal{Q}}_{\mu_I}^i, \mathcal{Q}_\rho^i \subseteq \Delta^{|\mathcal{C}_i|}$ . This will also prove relevant in our subsequent analysis of the case of comonotonic risk measures.

**3.2. The Comonotonic Case.** The results introduced in Section 3 and Section 3.1 become more specific when the risk measures  $\mathcal{Q}_I$  and  $\mathcal{Q}_C$  are further restricted to be comonotonic. We discuss a model with  $T = 2$ , but the approach and results readily extend to a finite number of time periods, a case which we revisit in Section 3.4.

We start by characterizing  $\mu_I$ , with its set of representing measures  $\mathcal{Q}_I$  and its down closure  $\text{sub}(\mathcal{Q}_I)$ . The central result here, formalized in the next proposition, is the identification of  $\mathcal{Q}_I$  with the base polytope corresponding to a particular Choquet capacity  $c$ . This analogy proves very useful in our analysis, since it allows stating all properties of  $\mathcal{Q}_I$  by employing known results

for base polytopes of polymatroid rank functions<sup>7</sup>, a concept studied extensively in combinatorial optimization (see Section B of the Appendix for all the results relevant to our treatment, and [27, 53] for a comprehensive review).

**PROPOSITION 5.** *Consider a naïve dynamic comonotonic risk measure  $\mu_I : \mathcal{X}_2 \rightarrow \mathbb{R}$ , with  $\mu_I(Y) \stackrel{\text{def}}{=} \max_{\mathbf{q} \in \mathcal{Q}_I} \mathbf{q}^T \mathbf{Y}, \forall Y \in \mathcal{X}_2$ . Then, there exists a Choquet capacity  $c : 2^{|\Omega_2|} \rightarrow \mathbb{R}$  such that*

1. *The set of measures  $\mathcal{Q}_I$  is given by the base polytope corresponding to  $c$ , i.e.,*

$$\mathcal{Q}_I \equiv \mathcal{B}_c \stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \mathbb{R}^{|\Omega_2|} : \mathbf{q}(S) \leq c(S), \forall S \subseteq \Omega_2, \mathbf{q}(\Omega_2) = c(\Omega_2) \right\}. \quad (14)$$

2. *The down-monotone closure of  $\mathcal{Q}_I$  is given by the polymatroid corresponding to  $c$ , i.e.,*

$$\text{sub}(\mathcal{Q}_I) \equiv \mathcal{P}_c \stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \mathbb{R}_+^{|\Omega_2|} : \mathbf{q}(S) \leq c(S), \forall S \subseteq \Omega_2 \right\}. \quad (15)$$

*Proof.* By Theorem 1 for comonotonic risk measures, there exists a Choquet capacity  $c$  such that  $\mathcal{Q}_I = \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \mathbf{q}(S) \leq c(S), \forall S \subseteq \Omega_2 \right\}$ . Since  $c(\Omega_2) = 1$ , this set can be rewritten equivalently as the base polytope corresponding to  $c$  (also refer to Corollary 7 of the Appendix for the argument that  $\mathcal{B}_c \subset \mathbb{R}_+^{|\Omega_2|}$ ). For the second claim, we can invoke a classical result in combinatorial optimization, that the downward monotone closure of the base polytope  $\mathcal{B}_c$  is exactly given by the *polymatroid* corresponding to the rank function  $c$ , i.e.,  $\mathcal{P}_c$  (see Theorem 12 in Section B).  $\square$

In particular, both sets  $\mathcal{Q}_I$  and  $\text{sub}(\mathcal{Q}_I)$  are polytopes contained in the non-negative orthant, and generally described by exponentially many inequalities, one for each subset of the ground set  $\Omega_2$ . However, evaluating the risk measure  $\mu_I$  for a given  $Y \in \mathcal{X}_2$  can be done in time polynomial in  $|\Omega_2|$ , by a simple Greedy procedure (see Theorem 13 in the Appendix or Lemma 4.92 in [25]).

In view of the results in Section 3.1, one may also seek a characterization of the tightest upper bound to  $\mu_I$ , i.e.,  $\widehat{\mu}_I$ , or of its set of representing measures  $\widehat{\mathcal{Q}}_{\mu_I}$ . Unfortunately, this seems quite difficult for general Choquet capacities  $c$  – a particular case when it is possible is when  $\mu_I$  is given by AVaR $_{\varepsilon}$ , a case discussed in our companion paper Huang et al. [32]. However, the result in Lemma 2 nonetheless proves useful for several of the results in this section.

The following result provides a characterization for the time-consistent and comonotonic risk measure  $\mu_C = \mu_1 \circ \mu_2$  as a *coherent* risk measure, by describing its set of representing measures  $\mathcal{Q}_C$  and its down-monotone closure  $\text{sub}(\mathcal{Q}_C)$ .

**PROPOSITION 6.** *Consider a two-period consistent, comonotonic risk measure  $\mu_C(Y) = \mu_1 \circ \mu_2$ , where  $\mu_t : \mathcal{X}_t \rightarrow \mathcal{X}_{t-1}$ . Then,*

1. *There exists  $\mathcal{Q}_C \subseteq \Delta^{|\Omega_2|}$  such that  $\mu_C(Y) \stackrel{\text{def}}{=} \max_{\mathbf{q} \in \mathcal{Q}_C} \mathbf{q}^T \mathbf{Y}, \forall Y \in \mathcal{X}_2$ .*
2. *The set of measures  $\mathcal{Q}_C$  is given by*

$$\begin{aligned} \mathcal{Q}_C &\stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \exists \mathbf{p} \in \Delta^{|\Omega_1|}, \begin{array}{l} \mathbf{p}(S) \leq c_1(S), \forall S \subseteq \Omega_1 \\ \mathbf{q}(U) \leq p_i \cdot c_{2|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1 \end{array} \right\} \\ &\equiv \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \exists \mathbf{p} \in \mathcal{B}_{c_1} : \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{B}_{p_i \cdot c_{2|i}}, \forall i \in \Omega_1 \right\}, \end{aligned}$$

where  $c_1 : 2^{|\Omega_1|} \rightarrow \mathbb{R}$  and  $c_{2|i} : 2^{|\mathcal{C}_i|} \rightarrow \mathbb{R}, \forall i \in \Omega_1$  are Choquet capacities, and  $\mathcal{B}_{c_1}, \mathcal{B}_{c_{2|i}}$  are the base polytopes corresponding to  $c_1$  and  $c_{2|i}$ , respectively.

3. *The downward monotone closure of  $\mathcal{Q}_C$  is given by*

$$\begin{aligned} \text{sub}(\mathcal{Q}_C) &\stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \mathbb{R}_+^{|\Omega_2|} : \exists \mathbf{p} \in \mathbb{R}_+^{|\Omega_1|}, \begin{array}{l} \mathbf{p}(S) \leq c_1(S), \forall S \subseteq \Omega_1, \\ \mathbf{q}(U) \leq p_i \cdot c_{2|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1 \end{array} \right\} \\ &= \left\{ \mathbf{q} \in \mathbb{R}_+^{|\Omega_2|} : \exists \mathbf{p} \in \mathcal{P}_{c_1} : \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{P}_{p_i \cdot c_{2|i}}, \forall i \in \Omega_1 \right\} \end{aligned}$$

where  $\mathcal{P}_{c_1}$  and  $\mathcal{P}_{p_i \cdot c_{2|i}}$  are the polymatroids associated with  $c_1$  and  $p_i \cdot c_{2|i}$ , respectively.

<sup>7</sup>We note that, with the exception of the normalization requirement  $c(\Omega) = 1$  that is unimportant for analyzing fundamental structural properties, the definition of a *Choquet capacity* is identical to that of a *rank function* of a *polymatroid* [27, Chapter 2]. Therefore, we use the two names interchangeably throughout the current paper.

*Proof.* The proof is technical, and involves a repeated application of ideas similar to those in the proof of Proposition 5. Therefore, we relegate it to Section C of the Appendix.  $\square$

As expected, the set of product measures  $\mathcal{Q}_C$  and its down-monotone closure  $\text{sub}(\mathcal{Q}_C)$  have a more complicated structure than  $\mathcal{Q}_I$  and  $\text{sub}(\mathcal{Q}_I)$ , respectively. However, they remain polyhedral sets, characterized by the base polytopes and polymatroids associated with particular Choquet capacities  $c_1$  and  $c_{2|i}$ . The inequality descriptions of  $\mathcal{Q}_C$  and  $\text{sub}(\mathcal{Q}_C)$  involve exponentially many constraints, but evaluating  $\mu_C(Y)$  at a given  $Y \in \mathcal{X}_2$  can still be done in time polynomial in  $|\Omega_2|$ , by using the Greedy procedure suggested in Theorem 13 in a recursive manner.

Because  $\mathcal{Q}_I$  and  $\mathcal{Q}_C$  are polytopes, they can also be described in terms of their extreme points. The description of the vertices of polymatroids and base polytopes has been studied extensively in combinatorial optimization (see Theorem 14 in the Appendix or [53] for details). Here, we apply the result for the case of  $\mathcal{Q}_I$ , and extend it to the special structure of the set  $\mathcal{Q}_C$ .

**PROPOSITION 7.** *Consider two risk measures  $\mu_I$  and  $\mu_C$ , as given by Proposition 5 and Proposition 6. Then,*

1. *The extreme points of  $\mathcal{Q}_I$  are given by*

$$q_{\sigma(i)} = c(\cup_{k=1}^i \sigma(k)) - c(\cup_{k=1}^{i-1} \sigma(k)), i \in [1, |\Omega_2|],$$

where  $\sigma \in \Pi(\Omega_2)$  is any permutation of the elements of  $\Omega_2$ .

2. *The extreme points of  $\mathcal{Q}_C$  are given by*

$$q_{\sigma_\ell(i)} = \left[ c_1(\cup_{k=1}^\ell \pi(k)) - c_1(\cup_{k=1}^{\ell-1} \pi(k)) \right] \cdot \left[ c_{2|\ell}(\cup_{k=1}^i \sigma_\ell(k)) - c_{2|\ell}(\cup_{k=1}^{i-1} \sigma_\ell(k)) \right], \forall i \in [1, |\mathcal{C}_\ell|], \forall \ell \in \Omega_1,$$

where  $\pi \in \Pi(\Omega_1)$  is any permutation of the elements of  $\Omega_1$ , and  $\sigma_\ell \in \Pi(\mathcal{C}_\ell)$  is any permutation of the elements of  $\mathcal{C}_\ell$  (for each  $\ell \in \Omega_1$ ).

*Proof.* Part (1) follows directly from the well-known characterization of the extreme points of an extended polymatroid, summarized in Theorem 14.

Part (2) follows by a repeated application of Theorem 14 to both  $\mathbf{p}$  and  $\mathbf{q}$  in the description of  $\mathcal{Q}_C$  of Proposition 6. In particular, any value of  $\mathbf{p}$  can be expressed as a convex combination of the extreme points  $\mathbf{p}^\pi$  of  $\mathcal{B}_{c_1}$  such that  $\mathbf{p} = \sum_{\pi \in \Pi(\Omega_1)} \lambda_\pi \mathbf{p}^\pi$  for appropriate  $\{\lambda_\pi\}_{\pi \in \Pi(\Omega_1)}$ . Now, for each  $\ell \in \Omega_1$  the value  $\mathbf{q}|_{\mathcal{C}_\ell} \in \mathcal{P}_{p_\ell \cdot c_{2|\ell}}$  can be similarly expressed as a convex combination of the extreme points  $\mathbf{q}_\ell^\sigma$  for an appropriate set of convex weights  $\{\xi_\sigma\}_{\sigma \in \Pi(\mathcal{C}_\ell)}$ , such that

$$\mathbf{q}|_{\mathcal{C}_\ell} = p_\ell \cdot \sum_{\sigma \in \Pi(\mathcal{C}_\ell)} \xi_\sigma \mathbf{q}_\ell^\sigma = \sum_{\pi \in \Pi(\Omega_1)} \sum_{\sigma \in \Pi(\mathcal{C}_\ell)} \lambda_\pi \xi_\sigma p_\ell^\pi \mathbf{q}_\ell^\sigma = \sum_{\pi \in \Pi(\Omega_1)} \sum_{\sigma \in \Pi(\mathcal{C}_\ell)} \chi_{\pi, \sigma} p_\ell^\pi \mathbf{q}_\ell^\sigma.$$

The proposition then follows directly from the fact that  $\chi_{\pi, \sigma}$  are themselves convex combination coefficients, and  $p_\ell^\pi \mathbf{q}_\ell^\sigma$  are extreme points.  $\square$

**3.3. Computing the Optimal Bounds  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$ .** With the representations provided above, we now derive our main technical result, establishing a method for computing the tightest multiplicative bounds for a pair of consistent and inconsistent comonotonic risk measures. The following theorem summarizes the result.

**THEOREM 4.** *For any pair of risk measures  $\mu_I$  and  $\mu_C$  as introduced in Section 3.2,*

$$\alpha_{\mu_C, \mu_I}^* = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \frac{\sum_{i \in S} \max_{U \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U)}{c_{2|i}(U)}}{c_1(S)} \quad (16)$$

$$\alpha_{\mu_I, \mu_C}^* = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_C)} \max_{S \subseteq \Omega_2} \frac{\mathbf{q}(S)}{c(S)}. \quad (17)$$

Furthermore, the value for  $\alpha_{\mu_C, \mu_I}^*$  remains the same if the outer maximization over  $\mathbf{q}$  is done over  $\mathcal{Q}_I, \text{ext}(\mathcal{Q}_I)$  or  $\text{ext}(\text{sub}(\mathcal{Q}_I))$ , and corresponding statements hold for  $\alpha_{\mu_I, \mu_C}^*$ .



*Proof.* To prove the first result, recall from Proposition 3 that for any  $Y \geq 0$ ,

$$\mu_I(Y) \leq \alpha \cdot \mu_C(Y) \Leftrightarrow \text{sub}(\mathcal{Q}_I) \subseteq \alpha \cdot \text{sub}(\mathcal{Q}_C).$$

Consider an arbitrary  $\mathbf{q} \in \text{sub}(\mathcal{Q}_I)$ . Any feasible scaling  $\alpha > 0$  must satisfy that  $\frac{1}{\alpha}\mathbf{q} \in \text{sub}(\mathcal{Q}_C)$ . Using the representation for  $\text{sub}(\mathcal{Q}_C)$  in Proposition 6, this condition yields

$$\frac{1}{\alpha}\mathbf{q} \in \text{sub}(\mathcal{Q}_C) \Leftrightarrow \exists \mathbf{p} \in \mathbb{R}_+^{|\Omega_1|} : \begin{cases} \mathbf{p}(S) \leq c_1(S), \forall S \subseteq \Omega_1 \\ \frac{1}{\alpha}\mathbf{q}(U) \leq p_i \cdot c_{2|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{cases}$$

The second set of constraints implies that any feasible  $\mathbf{p}$  satisfies  $p_i \geq \frac{1}{\alpha} \max_{U \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U)}{c_{2|i}(U)}$ ,  $\forall i \in \Omega_1$ . Corroborated with the first set of constraints, this yields

$$\begin{aligned} \frac{1}{\alpha} \sum_{i \in S} \max_{U \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U)}{c_{2|i}(U)} &\leq \sum_{i \in S} p_i \leq c_1(S), \forall S \subseteq \Omega_1 \Leftrightarrow \\ \alpha &\geq \max_{S \subseteq \Omega_1} \frac{\sum_{i \in S} \max_{U \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U)}{c_{2|i}(U)}}{c_1(S)}. \end{aligned}$$

Since this must be true for any  $\mathbf{q} \in \text{sub}(\mathcal{Q}_I)$ , the smallest possible  $\alpha$  is given by maximizing the expression above over  $\mathbf{q} \in \text{sub}(\mathcal{Q}_I)$ , which leads to the result (16).

The expression for  $\alpha_{\mu_I, \mu_C}^*$  is a direct application of the second part of Theorem 3, by identifying  $\text{sub}(\mathcal{Q}_1)$  with  $\text{sub}(\mathcal{Q}_I)$  and using the compact representation for  $\text{sub}(\mathcal{Q}_I)$  from Proposition 5.

The claim concerning the alternative sets follows by recognizing that the function maximized is always nondecreasing in the components of  $\mathbf{q}$ , so that  $\text{sub}(\mathcal{Q})$  can be replaced with  $\mathcal{Q}$ , and it is also convex in  $\mathbf{q}$ , hence reaching its maximum at the extreme points of the feasible set.  $\square$

From Theorem 4, it can readily be seen that, when  $\mu_C \leq \mu_I$ , the optimal  $\alpha_{\mu_C, \mu_I}^*$  will always be at least 1, and can be  $+\infty$  whenever the dimension of the polytope  $\mathcal{Q}_C$  is strictly smaller than that of  $\mathcal{Q}_I$ . Similarly, when  $\mu_I \leq \mu_C$ , the optimal  $\alpha_{\mu_I, \mu_C}^*$  is always at least 1, and can be  $+\infty$  when the dimension of the polytope  $\mathcal{Q}_I$  is smaller than  $\mathcal{Q}_C$ . To avoid the cases of unbounded optimal scaling factors, one can make the following assumption about the Choquet capacities.

**ASSUMPTION 1 (Relevance).** *The Choquet capacities  $c, c_1, c_{2|i}$  appearing in the representations for  $\mu_I$  and  $\mu_C$  (Proposition 5 and Proposition 6) satisfy the properties*

$$\begin{aligned} c(\{k\}) &> 0, \forall k \in \Omega_2 \\ c_1(\{i\}) &> 0, \forall i \in \Omega_1 \\ c_{2|i}(\{j\}) &> 0, \forall i \in \Omega_1, \forall j \in \mathcal{C}_i. \end{aligned}$$

This ensures that both risk measures consider all possible outcomes in the scenario tree, and is in line with the original requirement of *relevance* in [5], which states that, for any random cost  $Y$  such that  $Y \geq 0$  and  $Y \neq 0$ , any risk measure  $\mu$  should satisfy  $\mu(Y) > 0$ . In this case, the polytopes  $\mathcal{Q}_I$  and  $\mathcal{Q}_C$  are both full-dimensional (see [8] and Appendix A), which leads to finite minimal scalings.

As suggested in our general exposition at the beginning of Section 3, determining the optimal scaling factors  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  also leads to direct conditions for determining whether  $\mu_C$  lower bounds  $\mu_I$  or viceversa. The following corollary states these in terms of optimization problems.

**COROLLARY 3.** *For any pair of risk measures  $\mu_I$  and  $\mu_C$  as introduced in Section 3.2,*

1. *The inequality  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if*

$$\max_{\mathbf{q} \in \text{ext}(\mathcal{Q}_C)} \max_{S \subseteq \Omega_2} [\mathbf{q}(S) - c(S)] \leq 0.$$

2. The inequality  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if

$$\max_{q \in \text{ext}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \left[ \sum_{i \in S} \max_{U \subseteq \mathcal{C}_i} \frac{q(U)}{c_{2|i}(U)} - c_1(S) \right] \leq 0.$$

*Proof.* The proof follows from Corollary 2, by recognizing that the condition  $\mu_C \leq \mu_I$  is equivalent to setting  $\alpha_{\mu_I, \mu_C}^* \leq 1$  (and similarly for the reverse inequality and  $\alpha_{\mu_C, \mu_I}^*$ ). The formulas in Theorem 4 then immediately yield the desired conclusions.  $\square$

The results in Corollary 3 are stated in terms of non-trivial optimization problems. It is also possible to write out the conditions in a combinatorial fashion, using the analytical description of the extreme points of  $\mathcal{Q}_C$  and  $\mathcal{Q}_I$ , as summarized in the following corollary.

**COROLLARY 4.** *For any pair of risk measures  $\mu_I$  and  $\mu_C$  as introduced in Section 3.2,*

1. The inequality  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if

$$\sum_{j=1}^{|\Omega_1|} \left[ c_1(\cup_{k=1}^j s_k) - c_1(\cup_{k=1}^{j-1} s_k) \right] \cdot c_{2|s_j}(U_{s_j}) \leq c(\cup_{i \in \Omega_1} U_i),$$

where  $(s_1, \dots, s_{|\Omega_1|})$  denotes any permutation of the elements of  $\Omega_1$ , and  $U_i \subseteq \mathcal{C}_i$  for any  $i \in \Omega_1$ .

2. The inequality  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if

$$\begin{aligned} c(\cup_{i \in S} \mathcal{C}_i) &\leq c_1(S), \quad \forall S \subseteq \Omega_1, \\ \frac{c(U)}{c(U) + 1 - c(\Omega_2 \setminus \mathcal{C}_i \cup U)} &\leq c_{2,i}(U), \quad \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

*Proof.* The proof is slightly technical, so we defer it to Section C of the Appendix.  $\square$

The above conditions are explicit, and can always be checked when oracles are available for evaluating the relevant Choquet capacities. The main shortcoming of that approach is that the number of conditions to test is generally exponential in the size of the problem, even for a fixed  $T$ :  $\mathcal{O}((|\Omega_1|!) \cdot 2^{|\Omega_2|})$  for  $\mu_C \leq \mu_I$ , and  $\mathcal{O}(|A| \cdot 2^{\max_{i \in A} |\mathcal{C}_i|})$  for  $\mu_I \leq \mu_C$ , respectively, where  $A \stackrel{\text{def}}{=} \cup_{t \in [0, T-1]} \Omega_t$ . However, under additional assumptions on the Choquet capacities or the risk measures, it is possible to derive particularly simple polynomially-sized tests. We refer the interested reader to the discussion in Section 4.1 and the example in Section 4.3.

We note that the reason the conditions for  $\mu_I \leq \mu_C$  take a decoupled form and result in a smaller overall number of inequalities is directly related to the results of Lemma 2, which argues that testing  $\mu_I \leq \mu_C$  can be done by separately examining conditions at each node of the scenario tree.

**3.4. Multi-stage Extensions.** Although we focused our discussion thus far on a setting with  $T = 2$ , the ideas can be readily extended to an arbitrary, finite number of periods. We briefly outline the most relevant results in this section, but omit including the proofs, which are completely analogous to those for  $T = 2$ .

In a setting with general  $T$ , our goal is to compare a comonotonic  $\mu_I$  with a time-consistent, comonotonic  $\mu_C \stackrel{\text{def}}{=} \mu_1 \circ \mu_2 \circ \dots \circ \mu_T$ . The former is exactly characterized by Proposition 5, while the representation for the latter can be summarized in the following extension of Proposition 6.

**PROPOSITION 8.** *Consider a time-consistent, comonotonic risk measure  $\mu_C$ . Then,*

1. There exists  $\mathcal{Q}_C \subseteq \Delta^{|\Omega_T|}$  such that  $\mu_C(Y) \stackrel{\text{def}}{=} \max_{q \in \mathcal{Q}_C} q^T Y$ ,  $\forall Y \in \mathcal{X}_2$ .

2. The set of measures  $\mathcal{Q}_C$  is given by

$$\begin{aligned} \mathcal{Q}_C &\stackrel{\text{def}}{=} \left\{ \mathbf{p}_T \in \Delta^{|\Omega_T|} : \exists \{ \mathbf{p}_t \in \Delta^{|\Omega_t|} \}_{t \in [1, T-1]}, \mathbf{p}_t(U) \leq \mathbf{p}_{t-1}(\{i\}) \cdot c_{t|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_{t-1}, \forall t \in [1, T] \right\} \\ &\equiv \left\{ \mathbf{p}_T \in \Delta^{|\Omega_T|} : \exists \{ \mathbf{p}_t \in \Delta^{|\Omega_t|} \}_{t \in [1, T-1]}, \mathbf{p}_t|_{\mathcal{C}_i} \in \mathcal{B}_{\mathbf{p}_{t-1}(\{i\}) \cdot c_{t|i}}, \forall i \in \Omega_{t-1}, \forall t \in [1, T] \right\}, \end{aligned} \quad (18)$$

where  $c_{t|i} : 2^{\mathcal{C}_i} \rightarrow \mathbb{R}$  are Choquet capacities with corresponding base polytopes  $\mathcal{B}_{c_{t|i}}$ , for every  $t \in [1, T]$  and for every  $i \in \Omega_{t-1}$ .

3. The downward monotone closure  $\text{sub}(\mathcal{Q}_C)$  of  $\mathcal{Q}_C$  is obtained by replacing  $\Delta^{|\Omega_t|}$  with  $\mathbb{R}_+^{|\Omega_t|}$  and  $\mathcal{B}_{\mathbf{p}_{t-1}(\{i\}) \cdot c_{t|i}}$  with the polymatroid  $\mathcal{P}_{\mathbf{p}_{t-1}(\{i\}) \cdot c_{t|i}}$  in equation (18).

The proof exactly parallels that of Proposition 6, and is omitted due to space considerations. With this result, we can now extend our main characterization in Theorem 4 for the optimal multiplicative factors to a multi-period setting, as follows.

**THEOREM 5.** *For any comonotonic measure  $\mu_I$  and time-consistent comonotonic measure  $\mu_C$ ,*

$$\alpha_{\mu_C, \mu_I}^* = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \frac{\sum_{i \in S} z_1(i, \mathbf{q})}{c_1(S)}, \quad (19)$$

where  $\mathbf{z}_T(i, \mathbf{q}) \stackrel{\text{def}}{=} q_i$ ,  $\forall i \in \Omega_T$ , and  $z_t(i, \mathbf{q}) \stackrel{\text{def}}{=} \max_{U \subseteq \mathcal{O}_i} \frac{\sum_{i \in U} z_{t+1}(i, \mathbf{q})}{c_{t+1|U}(U)}$ ,  $\forall t \in [1, T-1]$ ,  $\forall i \in \Omega_t$ . Also,

$$\alpha_{\mu_I, \mu_C}^* = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_C)} \max_{S \subseteq \Omega_T} \frac{\mathbf{q}(S)}{c(S)}. \quad (20)$$

Furthermore, the value for  $\alpha_{\mu_C, \mu_I}^*$  would remain the same if the outer maximization were taken over  $\mathcal{Q}_I, \text{ext}(\mathcal{Q}_I)$  or  $\text{ext}(\text{sub}(\mathcal{Q}_I))$ . Corresponding statements hold for  $\alpha_{\mu_I, \mu_C}^*$ .

The proof follows analogously to that of Theorem 4, by using the expressions for  $\text{sub}(\mathcal{Q}_I)$  and  $\text{sub}(\mathcal{Q}_C)$  provided by Proposition 5 and Proposition 8, respectively, to analyze the conditions  $\text{sub}(\mathcal{Q}_I) \subseteq \alpha \cdot \text{sub}(\mathcal{Q}_C)$  or vice-versa. We omit it for brevity.

By comparing (22) and (20) with their two-period analogues in (16) and (17), respectively, it is interesting to note that the complexity of the formulation for  $\alpha_{\mu_I, \mu_C}^*$  remains the same, while the optimization problems yielding  $\alpha_{\mu_C, \mu_I}^*$  get considerably more intricate. Section 4 contains a detailed analysis of the computational complexity surrounding these problems.

For completeness, we remark that direct multi-period counterparts for Corollary 3 and Corollary 4 can be obtained, by recognizing that  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y$  is equivalent to  $\alpha_{\mu_I, \mu_C}^* \leq 1$ , and by using the results in Theorem 5 and Lemma 2 to simplify the latter conditions. We do not include these extensions due to space considerations.

**4. Discussion of the Results.** In view of the results in the previous section, several natural questions emerge. What is the computational complexity of determining the optimal scaling factors  $\alpha_{\mu_I, \mu_C}^*$  and  $\alpha_{\mu_C, \mu_I}^*$  for coherent/comonotonic risk measures? If this is generally hard, are there special cases that are easy, i.e., admitting polynomial-time algorithms? What examples of time-consistent risk measures can be derived starting with a given  $\mu_I$ , and how closely do they approximate the original measure?

The goal of the present section is to address these questions in detail. As we argue in Section 4.1, computing the scaling factors is hard even when restricting attention to *distortion* risk measures – a proper subclass of comonotonic measures. However, several relevant cases are nonetheless tractable. Section 4.2 introduces examples obtained by composing  $\mu_I$  with the conditional expectation operator “ $\mathbb{E}$ ” or conditional worst-case operator “ $\max$ ”, and compares them in terms of their approximation strength. Section 4.3 then summarizes the case when  $\mu_I$  and  $\mu_C$  correspond to the AVaR risk measure, and shows how many of the results drastically simplify.

**4.1. Computational Complexity.** As argued in Section 3, computing the optimal scaling factors  $\alpha_{\mu_I, \mu_C}^*$  and  $\alpha_{\mu_C, \mu_I}^*$  entails solving the optimization problems in (16) and (17). We now show that this is NP hard even for a problem with  $T = 1$ , and even when only examining *distortion* risk measures. We use a reduction from the SUBSET-SUM problem, which is NP hard [18] and is defined as follows.

DEFINITION 6 (SUBSET-SUM). Given a set of integers  $\{k_1, k_2, \dots, k_m\}$ , is there a subset that sums to  $s$ ?

This following result is instrumental in showing the complexity of computing the optimal scalings  $\alpha_{\mu_I, \mu_C}^*$  and  $\alpha_{\mu_C, \mu_I}^*$ .

THEOREM 6. Consider two arbitrary distortion risk measures  $\mu_{1,2} : \mathcal{X}_1 \rightarrow \mathbb{R}$ . Then, it is NP-hard to decide if  $\alpha_{\mu_2, \mu_1}^* \geq \gamma$ , for any  $\gamma \geq 0$ . The problem remains NP-hard even when  $\mu_2(Y) \leq \mu_1(Y)$ , for all  $Y \in \mathcal{X}_1$  ( $Y \geq 0$ ).

*Proof.* We use the representation of distortion risk measures to show the reduction from the SUBSET-SUM problem. By the representation Theorem 1 written for the specific case of distortion measures yields,  $\mu_i(Y) = \max_{\mathbf{q} \in \mathcal{Q}_i} \mathbf{q}^T \mathbf{Y}$ , where

$$\mathcal{Q}_i = \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \mathbf{q}(S) \leq c_i(S), \forall S \subseteq \Omega_1 \right\}, \quad \forall i \in \{1, 2\},$$

and  $c_i(S) = \Psi_i(\mathbb{P}(S))$ , where  $\Psi_i : [0, 1] \rightarrow [0, 1]$  are concave, increasing functions satisfying  $\Psi_i(0) = 0$ ,  $\Psi_i(1) = 1$ . Because both  $\text{sub}(\mathcal{Q}_1)$  and  $\text{sub}(\mathcal{Q}_2)$  are polymatroids and downward monotone, the second result in Theorem 4 can be further simplified to:

$$\alpha_{\mu_2, \mu_1}^* = \max_{S \subseteq \Omega_1} \frac{c_1(S)}{c_2(S)}. \quad (21)$$

Now, consider a SUBSET-SUM problem with values  $k_1, k_2, \dots, k_m$  and a value  $s$  such that  $1 \leq s < K$ , where  $K = \sum_{j=1}^m k_j$ . Construct the functions  $c_1$  and  $c_2$  as follows:

$$\mathbb{P}(s_i) = k_i/K \quad c_1(S) = \min \left\{ (\mathbb{P}(S) \cdot K) / s, 1 \right\} \quad c_2(S) = \min \left\{ c_1(S), \sqrt{\mathbb{P}(S)} \right\}$$

Since both  $c_1, c_2$  satisfy the conditions of distortion risk measures, any SUBSET-SUM problem can be reduced to the problem of computing the optimal scale of two distortion risk measures.

Now, the optimal value of (21) is upper bounded as:

$$\max_{S \subseteq \Omega_1} \frac{c_1(S)}{c_2(S)} \leq \sqrt{\frac{K}{s}}.$$

The maximum is achieved when there exists  $S$  such that  $\mathbb{P}(S) = s/K$ . To show this, consider  $c_1(S)/c_2(S)$  as a function of  $\mathbb{P}(S)$ . This function is: (1) non-decreasing on the interval  $[0, s/K]$  and non-increasing on the interval  $(s/K, 1]$ , (2) strictly greater than one for  $\mathbb{P}(S) = s/K$ , (3) equal to 1 for  $\mathbb{P}(S) \in \{0, 1\}$ , and (4) continuous. Therefore, the SUBSET-SUM problem has a subset that sums to  $s$  if and only if the optimal value of (21) is  $\sqrt{K/s}$ . Finally, the result also holds when  $\mu_2(Y) \leq \mu_1(Y)$ , since our choice already has  $c_2(S) \leq c_1(S)$  for all  $S \subseteq \Omega_1$ , which implies  $\mu_2(Y) \leq \mu_1(Y)$ .  $\square$

The complexity of computing the optimal scalings  $\alpha_{\mu_I, \mu_C}^*$  and  $\alpha_{\mu_C, \mu_I}^*$  readily follows as a direct corollary of Theorem 6.

COROLLARY 5. Under a fixed  $T \geq 1$  and for any given  $\gamma \geq 0$ , it is NP-complete to decide whether  $\alpha_{\mu_C, \mu_I}^* \geq \gamma$  for an arbitrary inconsistent distortion measure  $\mu_I$  and a consistent distortion measure  $\mu_C$ . The result remains true even when  $\mu_C$  and  $\mu_I$  are such that  $\mu_C(Y) \leq \mu_I(Y)$  for all  $Y \in \mathcal{X}_T$  ( $Y \geq 0$ ). Similarly, it is NP-complete to decide whether  $\alpha_{\mu_I, \mu_C}^* \geq \gamma$ , and the result remains true even when  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y \in \mathcal{X}_T$  ( $Y \geq 0$ ).

*Proof.* First, note that finding the scaling factors for any  $T > 1$  is at least as hard as for  $T = 1$ . This can be seen by setting  $|\Omega_t| = 1$  for all  $t \in [2, T - 1]$ . The NP-hardness then follows from Theorem 6 by setting  $\mu_2 = \mu_C$  and  $\mu_1 = \mu_I$ . The membership in NP follows by checking the inequality (16) for every extreme point  $\mathbf{q}$ , subset  $S$ , and the appropriate subsets  $U$ . The second result follows analogously.  $\square$

Corollary 5 argues that computing the optimal scaling factors for arbitrary distortion risk measures cannot be done in polynomial time. While the NP-hardness may be somewhat disappointing, solving the two optimization problems in Theorem 4 is nonetheless clearly preferable to simply examining all possible values of  $Y$ .

While the problem of computing the scaling factors is hard for general distortion measures, polynomial-time algorithms are possible when the representations of  $\mathcal{Q}_I, \mathcal{Q}_C$  or  $\text{sub}(\mathcal{Q}_I), \text{sub}(\mathcal{Q}_C)$  fall in the tractable cases discussed in Table 1 of Section 3.

In fact, some of the results of Table 1 can even be strengthened - one such case is when a vertex description for the polytope  $\mathcal{Q}_I$  is available, and problem (22) can be solved in time polynomial in  $|\Omega_T|$ , under oracle access to the Choquet capacities  $c_{t|i}$  yielding the measure  $\mu_C$ .

**LEMMA 3.** *If the polytope  $\mathcal{Q}_I$  is specified by a polynomial number of extreme points, then  $\alpha_{\mu_C, \mu_I}^*$  can be computed in time polynomial in  $|\Omega_T|$ .*

*Proof.* Consider the specialization of (22) for a fixed  $\mathbf{q} \in \mathcal{Q}_I$ :

$$\alpha_{\mu_C, \mu_I}^* = \max_{\mathbf{q} \in \text{sub}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \frac{z_1(U, \mathbf{q})}{c_1(S)}, \quad (22)$$

where  $z_T(i, \mathbf{q}) \stackrel{\text{def}}{=} q_i, \forall i \in \Omega_T$ , and  $z_t(i, \mathbf{q}) \stackrel{\text{def}}{=} \max_{U \subseteq \mathcal{C}_i} \frac{z_{t+1}(U, \mathbf{q})}{c_{t+1|i}(U)}, \forall t \in [1, T-1], \forall i \in \Omega_t$ .

Note that each value  $z_t(i, \mathbf{q})$  and also  $\alpha_{\mu_C, \mu_I}^*$  can be written as:

$$z_t(i, \mathbf{q}) = \max_{U \subseteq \mathcal{C}_i} \frac{z_{t+1}(U, \mathbf{q})}{c_{t+1|i}(U)} = \min \left\{ l \in \mathbb{R} : l \cdot c_{t+1|i}(U) - z_{t+1}(U, \mathbf{q}) \geq 0, \forall U \subseteq \mathcal{C}_i \right\}.$$

For any  $l$ , the constraint  $l \cdot c_{t+1|i}(U) - z_{t+1}(U, \mathbf{q}) \geq 0, \forall U \subseteq \mathcal{C}_i$  can be checked in polynomial time, since the set function on the left-hand side is submodular in  $U$ , and can be minimized with a polynomial number of function evaluations [53].  $\square$

The result above is slightly stronger than what Table 1 suggests, since the representation of  $\mathcal{Q}_C \equiv \mathcal{Q}_1$  can still be exponential both in terms of extreme points and vertices, as long as oracle access to  $c_{t|i}$  is available.

**4.2. Examples.** To see how our results can be used to examine the tightness of particular dynamically consistent risk measures, we now consider several constructions suggested in the literature. The starting point is typically a single *distortion* risk measure  $\mu_I : \mathcal{X}_2 \rightarrow \mathbb{R}$ , denoting the inconsistent evaluation. This is then composed with other suitable measures (for instance, with itself, with the conditional expectation and/or the conditional worst-case operator), to obtain time-consistent risk measures that are derived from  $\mu_I$ . The questions we would like to address here is which of these measures lower-bound or upper-bound the inconsistent evaluation  $\mu_I$ , and what can be said about the relative tightness of the various formulations.

In order to construct dynamically-consistent measures by composing  $\mu_I$ , we must first specify the conditional one-step risk mappings corresponding to  $\mu_I$ , formally denoted by  $\mu_I^1 : \mathcal{X}_1 \rightarrow \mathbb{R}$  and  $\mu_I^2 : \mathcal{X}_2 \rightarrow \mathcal{X}_1$ . When  $\mu_I$  is a *distortion* risk measure, this can be done in a natural way in terms of the corresponding concave distortion function. To this end, recall that, by the representation Theorem 1, any distortion measure  $\mu_I$  is uniquely specified by the concave function  $\Psi$  yielding its set of representing measures, through the Choquet capacity  $c(S) = \Psi(\mathbb{P}(S)), \forall S \subseteq \Omega_2$ . The conditional one-step risk mappings  $\mu_I^1$  and  $\mu_I^2 \equiv (\mu_I^{2|i})_{i \in \Omega_1}$  are then obtained by applying the same distortion function  $\Psi$  to suitable conditional probabilities. More precisely,  $\mu_I^1$  and  $\mu_I^{2|i}$  are the distortion risk measures corresponding to the Choquet capacities:

$$c_1 : 2^{\Omega_1} \rightarrow \mathbb{R}, c_1(S) = \Psi \left( \sum_{i \in S} \mathbb{P}(\mathcal{C}_i) \right), \forall S \subseteq \Omega_1$$

$$c_{2|i} : 2^{\mathcal{C}_i} \rightarrow \mathbb{R}, c_{2|i}(U_i) = \Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right), \forall U_i \subseteq \mathcal{C}_i, \forall i \in \Omega_1.$$

The conditional risk mappings  $\mu_I^1$  and  $\mu_I^2$  can be used to define dynamic time-consistent risk measures, either alone or by composition with other conditional risk mappings. In particular, all of the following dynamic time-consistent risk measures have been considered in the literature:

$$\mathbb{E} \circ \mu_I^2 \quad \mu_I^1 \circ \mathbb{E} \quad \mu_I^1 \circ \mu_I^2 \quad \mu_I^1 \circ \max \quad \max \circ \mu_I^2,$$

where  $\mathbb{E}$  denotes the conditional expectation operator, and  $\max$  is the conditional worst-case operator. Whenever the meaning is clear from context, we sometimes omit the time-subscript, and use shorthand notation such as  $\mathbb{E} \circ \mu_I$ ,  $\mu_I \circ \mathbb{E}$ ,  $\mu_I \circ \mu_I$ , etc., although we are formally referring to compositions with  $\mu_I^1$  and/or  $\mu_I^2$ .

**4.2.1. Time-Consistent Lower Bounds Derived From a Given  $\mu_I$ .** We begin by discussing two choices for lower-bounding consistent risk measures derived from  $\mu_I$ . The following proposition formally establishes the first relevant result.

**PROPOSITION 9.** *Consider any distortion risk measure  $\mu_I : \mathcal{X}_2 \rightarrow \mathbb{R}$ , and the time-consistent, comonotonic measures  $\mu_I \circ \mathbb{E}$  and  $\mathbb{E} \circ \mu_I$ . Then, for any cost  $Y \in \mathcal{X}_2$ ,*

$$(\mu_I \circ \mathbb{E})(Y) \leq \mu_I(Y) \quad \text{and} \quad (\mathbb{E} \circ \mu_I)(Y) \leq \mu_I(Y).$$

*Proof.* The proof entails directly checking the conditions in Corollary 4. A complete derivation is included in Section C of the Appendix.

This is not a surprising result, since the  $\mathbb{E}$  operator is known to be a uniform lower bound for any static coherent risk measure [25]. We confirm that the same remains true in dynamic settings, provided that the risk measure  $\mu_I$  is applied in a single time step, and conditional expectation operators are applied in other stages.

Since both  $\mu_I \circ \mathbb{E}$  and  $\mathbb{E} \circ \mu_I$  are lower bounds for  $\mu_I$ , a natural question is whether one provides a “better” approximation than the other. More precisely, the following are questions of interest:

1. For a given  $\mu_I$ , is it true that  $(\mu_I \circ \mathbb{E})(Y) \leq (\mathbb{E} \circ \mu_I)(Y)$ ,  $\forall Y \in \mathcal{X}_2$  (or vice-versa)?
2. Is it true that  $\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^* \geq \alpha_{\mathbb{E} \circ \mu_I, \mu_I}^*$  for any distortion measure  $\mu_I$  (or vice-versa)?

Clearly, a positive answer to the first question would provide a very strong sense of tightness of approximation. However, as the following example shows, neither inequality holds in general.

**EXAMPLE 2.** Consider a scenario tree with  $T = 2$ ,  $|\Omega_1| = 2$ ,  $|\mathcal{C}_i| = 2$ ,  $\forall i \in \Omega_1$ , under uniform reference measure. Introduce the following two random costs  $X, Y$  (specified as vectors in  $\mathbb{R}^{|\Omega_2|}$ ):

$$\begin{aligned} X|_{\mathcal{C}_1} &= M \cdot \mathbf{1}, & X|_{\mathcal{C}_2} &= \mathbf{0} \\ Y|_{\mathcal{C}_1} &= Y|_{\mathcal{C}_2} = [M, 0]^T. \end{aligned}$$

With  $M > 0$ , and  $\mu_I \equiv \text{AVaR}_{1/2}$ , it can be checked that  $(\mu_I \circ \mathbb{E})(X) = M > (\mathbb{E} \circ \mu_I)(X) = \frac{M}{2}$ , while  $(\mu_I \circ \mathbb{E})(Y) = \frac{M}{2} < (\mathbb{E} \circ \mu_I)(Y) = M$ .

Insofar as the second question is concerned, we note that it can always be answered for a *specific* distortion measure  $\mathcal{Q}_I$ , by calculating the optimal scalings, so that it really makes sense when posed for *all* risk measures. Unfortunately, our computational experiments show that counterexamples can be constructed for this claim, as well, and that any one of the scaling factors can be better than the other. However, it would be very interesting to characterize conditions (on the risk measures, the underlying probability space, or otherwise) under which a particular compositional form always results in a smaller scaling factor. The following result, which we prove in the Appendix, is a potential first step in this direction, suggesting that the two lower bounds can result in *equal* tightness of approximation in certain cases of interest.

**THEOREM 7.** *Consider a uniform scenario tree, i.e.,  $|\Omega_1| = N$ ,  $|\mathcal{C}_i| = N$ ,  $\forall i \in \Omega_1$ , under a uniform reference measure. Then, for any distortion risk measure  $\mu_I$ , we have*

$$\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^* = \alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* = N \cdot \max \left\{ \frac{\Psi(1/N^2)}{\Psi(1/N)}, \frac{\Psi(2/N^2)}{\Psi(2/N)}, \dots, \Psi(1/N) \right\}.$$

**4.2.2. Time-Consistent Upper Bounds Derived From a Given  $\mu_I$ .** In an analogous fashion to the previous discussion, one can ask what time-consistent *upper bounds* can be derived from a distortion measure  $\mu_I$ . In particular, a natural supposition, analogous to the results of Section 4.2.1, may be that  $\mu_I \circ \max$  and  $\max \circ \mu_I$  are upper bounds to  $\mu_I$ , since  $\max$  is the most conservative risk mapping possible [25]. The following result shows that, unlike in the lower bound setting, *only one* of the two composed measures is a valid upper bound.

**PROPOSITION 10.** *Consider any distortion risk measure  $\mu_I$ , and the time-consistent, comonotonic measures  $\mu_I \circ \max$  and  $\max \circ \mu_I$ , where  $\max$  denotes the conditional worst-case operator. Then:*

(i) *For any cost  $Y \in \mathcal{X}_2$ ,  $\mu_I(Y) \leq (\mu_I \circ \max)(Y)$ .*

(ii) *There exists a choice of  $\mu_I$  and of random costs  $Y_{1,2} \in \mathcal{X}_2$  such that  $(\max \circ \mu_I)(Y_1) < \mu_I(Y_1)$  and  $(\max \circ \mu_I)(Y_2) > \mu_I(Y_2)$ .*

*Proof.* The proof for Part (i) entails checking the conditions of Corollary 4. Since it is rather technical in nature, we leave it for Section C of the Appendix of the paper.

To show Part (ii), consider a uniform scenario tree with  $|\Omega_1| = |\mathcal{C}_i| = 2$ ,  $\forall i \in \Omega_1$ , and let the reference measure be  $\mathbb{P} = [0.1, 0.5, 0.2, 0.3]^T$ . For simplicity, assume the first two components of  $\mathbb{P}$  correspond to nodes in the same child. Then, for the risk measure  $\mu_I = \text{AVaR}_{1/2}$ , and the costs  $\mathbf{Y}_1 = [1, 0, 0, 0.4]^T$  and  $\mathbf{Y}_2 = [0, 0, 0, 1]^T$ , it can be checked that  $\mu_I(\mathbf{Y}_1) = 0.44 > (\max \circ \mu_I)(\mathbf{Y}_1) = 0.4$ , but  $\mu_I(\mathbf{Y}_2) = 0.3 < (\max \circ \mu_I)(\mathbf{Y}_2) = 1$ .  $\square$

The result in Proposition 10 also suggests that upper bounds to  $\mu_I$  can be derived by composing  $\mu_I$  with more conservative mappings in later time periods. This intuition is sharpened in Section 4.3 and our companion paper [32], which show that, when  $\mu_I = \text{AVaR}_\varepsilon$ , all upper bounds of the form  $\text{AVaR}_\varepsilon \circ \text{AVaR}_\gamma$  must have  $\gamma \leq \varepsilon$ , and, in many practical settings,  $\gamma = 0$ , i.e., worst-case as the second-stage evaluation.

Since  $\mu_I \circ \max$  is an upper bound for a given  $\mu_I$ , one can also turn to the question of comparing the resulting scaling factor  $\alpha_{\mu_I, \mu_I \circ \max}^*$  with the factors of the previous section, namely  $\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^*$  or  $\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^*$ . Our computational tests show that there is no general relation between these, even when the scenario tree and the reference measure are uniform, a claim due to the following result, whose proof is included in the paper's Appendix.

**PROPOSITION 11.** *Consider a uniform scenario tree, i.e.,  $|\Omega_1| = N$ ,  $|\mathcal{C}_i| = N$ ,  $\forall i \in \Omega_1$ , under a uniform reference measure. Then, for any distortion risk measure  $\mu_I$ , we have*

$$\alpha_{\mu_I, \mu_I \circ \max}^* = \max \left\{ \frac{\Psi(1/N)}{\Psi(1/N^2)}, \frac{\Psi(2/N)}{\Psi(2/N^2)}, \dots, \frac{1}{\Psi(1/N)} \right\}.$$

Corroborating this result with the expression in Theorem 7 for  $\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^*$ , one can readily find simple examples of distortions  $\Psi$  such that either the latter or the former scaling factor is smaller.

An opinion often held among practitioners, and informally argued in the literature [46, 44] is that composing a risk measure with itself would compound the losses, resulting in a larger evaluation of risk, i.e., that  $\mu_I \circ \mu_I$  should over-bound  $\mu_I$ . For instance, if  $\mu_I = \text{AVaR}$  – the case considered in [46] – the compositional measure corresponds to the so-called “iterated tail-CTE”, which takes tail conditional expectations of quantities that are already tail conditional expectations. We show by means of an example that this informal belief is actually *not* true, even in the case of AVaR.

EXAMPLE 3 (ITERATED AVaR). Consider a uniform scenario tree (i.e.,  $|\Omega_1| = |\mathcal{C}_i| = 4, \forall i \in \Omega_1$ ), and a uniform reference measure. Furthermore, consider the risk measure  $\mu_I \equiv \text{AVaR}_{3/4}$ , and the following two costs (specified as real vectors in  $\mathbb{R}^{|\Omega_2|}$ , with components split in the four sub-trees of stage  $T = 2$ ):

$$\begin{aligned} X|_{\mathcal{C}_1} = X|_{\mathcal{C}_2} = \mathbf{1}, \quad X|_{\mathcal{C}_3} = X|_{\mathcal{C}_4} = [1, 1, -M, -M]^T \\ Y|_{\mathcal{C}_1} = Y|_{\mathcal{C}_2} = Y|_{\mathcal{C}_3} = [1, 1, 1, -M]^T, \quad Y|_{\mathcal{C}_4} = -M \cdot \mathbf{1}. \end{aligned}$$

When  $M > -1$ , it can be readily checked<sup>8</sup> that  $\mu_I(X) = 1 > (\mu_I \circ \mu_I)(X) = \frac{8-M}{9}$ , while  $(\mu_I \circ \mu_I)(Y) = 1 > \mu_I(Y) = \frac{3-M}{4}$ .

The example shows that the iterated AVaR is *neither an upper nor a lower bound* to the static AVaR. We direct the interested reader to our companion paper [32], which is focused specifically on the AVaR case, and discusses the exact necessary and sufficient conditions for when one of the two dominates the other.

**4.2.3. The Tightest Possible Time-Consistent Upper-bound.** A natural time-consistent upper bound to a given  $\mu_I$  is the measure  $\widehat{\mu}_I$ , obtained by the rectangularization procedure in Proposition 4. It is the tightest possible coherent upper bound to  $\mu_I$ , both in a uniform and multiplicative-alpha sense. The main potential drawback in using  $\widehat{\mu}_I$  is that it may not satisfy additional axiomatic properties, and it typically bears no interpretation in terms of  $\mu_I$ . For instance, starting with a comonotonic  $\mu_I$  does *not* generally result in a comonotonic  $\widehat{\mu}_I$ , and  $\widehat{\mu}_I$  is usually not given by compositions of one-step risk measures that correspond to  $\mu_I$ . Determining conditions that guarantee the latter two properties is an interesting question, which we do not pursue further in the present paper. However, we note that this is possible in at least one case of practical interest: when  $\mu_I = \text{AVaR}_\varepsilon$ , one can show that  $\widehat{\mu}_I$  always corresponds to a composition of one-step AVaR measures, at appropriate levels – see our discussion in Section 4.3 and the detailed treatment in our companion paper [32].

For completeness, we also note that the ordering relation between the scaling factor  $\alpha_{\mu_I, \widehat{\mu}_I}^*$  and scalings  $\alpha_{\mu_C, \mu_I}^*$  derived from lower-bounding measures  $\mu_C$  is generally not obvious: our computational experiments suggest that either one could dominate the other. However, more can be said in particular settings, such as the case of AVaR, which we discuss next.

**4.3. The Case of AVaR.** In this section, we discuss how several of the results introduced throughout the paper can be considerably simplified when the risk measures in question correspond to AVaR. In particular, analytical expressions or polynomial-time procedures can be derived for computing  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  and for testing  $\mu_I(Y) \leq \mu_C(Y)$  or viceversa. Furthermore, one can consider designing the risk measures  $\mu_C$  that provide the tightest possible lower or upper approximations to a given  $\mu_I$ .

The case is discussed at length in our companion paper [32], to which we direct the interested reader for any technical details and proofs. Our goal for the remainder of the section is to outline the main results, and briefly discuss the implications.

To start, we consider a uniform scenario tree under uniform reference measure ( $|\mathcal{C}_i| = N, \forall i \in \cup_{t=0}^{T-1} \Omega_t$ , and  $\mathbb{P} = \frac{1}{N^T}$ ), and the following choice of risk measures:

$$\mu_I = \text{AVaR}_\varepsilon, \quad \varepsilon \in [1/N^T, 1] \quad (23a)$$

$$\mu_C = \text{AVaR}_{\varepsilon_1} \circ \text{AVaR}_{\varepsilon_2} \circ \dots \circ \text{AVaR}_{\varepsilon_T}, \quad \varepsilon_t \in [1/N, 1], \forall t \in [1, T]. \quad (23b)$$

<sup>8</sup> For the case of discrete probability measures, one has to be careful in defining  $\text{AVaR}_\varepsilon$ , since it is no longer exactly given by the conditional expectation of the loss exceeding  $\text{VaR}_\varepsilon$ . The precise concepts are presented and discussed at length in [41], which we follow here.



Note that the restriction on  $\varepsilon$  and  $\varepsilon_t$  is without loss of generality, since  $\text{AVaR}_\varepsilon$  with  $\varepsilon \leq \frac{1}{N^T}$  is identical to the worst-case risk measure, rendering the case  $\varepsilon \in [0, \frac{1}{N^T})$  analogous to  $\varepsilon = \frac{1}{N^T}$ .

In this setup, we can revisit our main results in Theorem 4, and provide the following expressions for the tightest factors  $\alpha_{\mu_C, \mu_I}^*$  and  $\alpha_{\mu_I, \mu_C}^*$  for the case  $T = 2$ .

**THEOREM 8.** *Consider a case  $T = 2$ , and the pair of risk measures in (23a) and (23b). Then,*

$$\alpha_{\mu_C, \mu_I}^* = \begin{cases} \max\{N\varepsilon_1, \frac{\varepsilon_1\varepsilon_2}{\varepsilon}, N\varepsilon_2\}, & \varepsilon \leq \frac{1}{N} \\ \max\{\frac{\varepsilon_1}{\varepsilon}, f(N, \varepsilon, \varepsilon_2)\}, & \varepsilon > \frac{1}{N}, \end{cases} \quad (24a)$$

$$\alpha_{\mu_I, \mu_C}^* = \max\left\{1, \frac{\varepsilon}{\varepsilon_1 \varepsilon_2}\right\}, \quad (24b)$$

where  $f(N, \varepsilon, \varepsilon_2)$  is an explicit analytical function. Furthermore, the result for  $\alpha_{\mu_I, \mu_C}^*$  remains true under an arbitrary scenario tree and reference measure  $\mathbb{P}$ .

Note that the above result has several immediate implications. First, it readily allows checking whether  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y$  (or vice-versa), since the latter conditions are equivalent to  $\alpha_{\mu_I, \mu_C}^* \leq 1$  (respectively,  $\alpha_{\mu_C, \mu_I}^* \leq 1$ ). This leads to the following simple tests.

**COROLLARY 6.** *Consider the pair of risk measures in (23a) and (23b). Then,*

1. *the inequality  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if*

$$\varepsilon_1 \varepsilon_2 \geq \varepsilon, \quad (25)$$

2. *the inequality  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if*

$$\varepsilon_1 \leq \max\left(\frac{1}{N}, \varepsilon\right) \quad \text{and} \quad \varepsilon_2 \leq \max\left(\frac{1}{N}, N\varepsilon - N + 1\right). \quad (26)$$

Furthermore, (25) remains true under an arbitrary scenario tree and reference measure  $\mathbb{P}$ .

The latter result confirms the observation in Example 3 that the iterated AVaR, i.e.,  $\mu_C = \text{AVaR}_\varepsilon \circ \text{AVaR}_\varepsilon$ , is generally neither an upper nor a lower bound to the inconsistent choice  $\mu_I = \text{AVaR}_\varepsilon$ . By (26),  $\varepsilon_1 = \varepsilon$  is always a feasible option, but one must take  $\varepsilon_2 \leq \max(1/N, N\varepsilon - N + 1)$ . In fact, as argued in [32], most relevant choices of  $\varepsilon$  would actually lead to taking  $\varepsilon_2 = 1/N$ , i.e., the worst-case operator in the second stage.

The analytical results above can also be used to *optimally design* the compositional risk measure  $\mu_C$  that is the tightest approximation to a given  $\mu_I = \text{AVaR}_\varepsilon$ . More precisely, one can characterize the choice of  $\mu_C$  (i.e., levels  $\varepsilon_{1,2}^{\text{LB}}$ ) that results in the smallest possible factor  $\alpha_{\mu_C, \mu_I}^*$  among all compositional AVaR that are lower bounds for  $\text{AVaR}_\varepsilon$ , and, similarly, the values  $\varepsilon_{1,2}^{\text{UB}}$  yielding the smallest possible  $\alpha_{\mu_I, \mu_C}^*$  among all upper-bounding compositional AVaRs. The optimal choices satisfy several interesting properties:

- for values of  $\varepsilon$  that are common in financial applications, i.e., satisfying  $\varepsilon \leq 1/N$  [35], the optimal  $\alpha_{\mu_C, \mu_I}^*$  is obtained by taking  $\varepsilon_1^{\text{LB}} = \varepsilon_2^{\text{LB}} = \sqrt{\varepsilon}$ , corresponding to an iterated AVaR measure.
- the optimal  $\alpha_{\mu_I, \mu_C}^*$  requires choosing  $\varepsilon_1^{\text{UB}} = \varepsilon$  and  $\varepsilon_2^{\text{UB}} = \max(1/N, N\varepsilon - N + 1)$ . Typical values of  $\varepsilon$  used in practice would entail  $\varepsilon_2^{\text{UB}} = 1/N$ , i.e., the worst-case scenario in the second stage.
- the optimally designed  $\alpha_{\mu_C, \mu_I}^*$  is *always* smaller than  $\alpha_{\mu_I, \mu_C}^*$ , i.e., for every  $\varepsilon$  and  $N$ , which suggests that starting with an under-estimating AVaR $_\varepsilon$  results in tighter dynamically consistent approximations for AVaR.

The results discussed in Theorem 8 for  $T = 2$  can also be (partially) extended to a case of an arbitrary  $T$ , which is summarized in the following claim.

**THEOREM 9.** *Consider an arbitrary  $T$ , and the pair of risk measures in (23a) and (23b). Then,*

1. *There is an algorithm that computes  $\alpha_{\mu_C, \mu_I}^*$  in time  $\mathcal{O}(N^{T^2})$ .*

2.  $\alpha_{\mu_I, \mu_C}^* = \max\left\{1, \frac{\varepsilon}{\prod_{t=1}^T \varepsilon_t}\right\}$ , and the expression remains valid for an arbitrary scenario tree and reference measure.

It is interesting to note that computing  $\alpha_{\mu_I, \mu_C}^*$ , and hence also testing  $\mu_C \leq \mu_I$ , remains as easy for general  $T$  as for  $T = 2$ : an analytical expression is available, which actually holds in considerably more general settings (arbitrary tree and reference measure). By contrast, computing  $\alpha_{\mu_C, \mu_I}^*$  and testing  $\mu_I \leq \mu_C$  now requires an algorithm that is polynomial only for a fixed  $T$ . In light of our earlier result, this suggests that, although starting with lower-bounds for  $\mu_I$  may lead to a tighter approximating  $\mu_C$ , the gain does not come for free, as the computation of the resulting  $\alpha_{\mu_C, \mu_I}^*$  is typically harder than that for  $\alpha_{\mu_I, \mu_C}^*$ .

In a multiperiod setting, the question of designing the tightest possible lower-bounding approximation  $\mu_C$  to a given  $\mu_I$  becomes harder – even computing one scaling factor  $\alpha_{\mu_C, \mu_I}^*$  requires a polynomial-time algorithm. By contrast, a complete characterization of the tightest upper-bound  $\mu_C$  is available! Quite surprisingly, it turns out that this choice exactly corresponds to the risk measure  $\hat{\mu}_I$  introduced by the construction in Proposition 4, by expanding the set of measures of  $\mu_I$ . This is summarized in the following result (for a proof, see [32]).

**THEOREM 10.** *Consider the risk measure  $\mu_I = \text{AVaR}_\varepsilon$ , under an arbitrary reference measure  $\mathbb{P}$ , and the construction for the risk measure  $\hat{\mu}_I$  characterized in Proposition 4. Then,  $\hat{\mu}_I = \text{AVaR}_1 \circ \text{AVaR}_2 \circ \dots \circ \text{AVaR}_T$ , where  $\text{AVaR}_t = (\hat{\mu}_i)_{i \in \Omega_{t-1}}$ , and*

$$\forall i \in \Omega_{t-1}, \quad \hat{\mu}_i = \begin{cases} \max, & \text{if } \mathbb{P}(\mathcal{D}_i) \leq 1 - \varepsilon \\ \text{AVaR}_{\gamma_i}, & \text{otherwise.} \end{cases}$$

Here,  $\gamma_i = \frac{\mathbb{P}(\mathcal{D}_i) - 1 + \varepsilon}{\mathbb{P}(\mathcal{D}_i)}$ , and  $\text{AVaR}_{\gamma_i}$  is computed under the conditional probability induced by  $\mathbb{P}$ , i.e.,  $\left(\frac{\mathbb{P}(\mathcal{D}_j)}{\mathbb{P}(\mathcal{D}_i)}\right)_{j \in \mathcal{C}_i}$ .

This result, which holds under any reference measure  $\mathbb{P}$ , suggests that starting with  $\mu_I = \text{AVaR}_\varepsilon$  and expanding its set of representing probability measures until it becomes rectangular exactly results in a risk measure  $\hat{\mu}_I$  that is a composition of one-step AVaRs. These one-step AVaRs are computed under levels  $\gamma_i$  that can be different at each node  $i$  in the tree, and under the natural conditional probability induced by the reference measure  $\mathbb{P}$ .

There are several immediate implications. First, since  $\hat{\mu}_I$  is the tightest possible coherent upper-bound for any given coherent  $\mu_I$  (see Lemma 2), this implies that the tightest possible choice for a compositional AVaR that upper bounds a given  $\text{AVaR}_\varepsilon$  is exactly  $\hat{\text{AVaR}}_\varepsilon$ . In a different sense, this also provides an instance when starting with a comonotonic (in fact, distortion) risk measure  $\mu_I$  results in a comonotonic (distortion) risk measure  $\hat{\mu}_I$ , which furthermore belongs to the same class as  $\mu_I$ .

Lastly, the theorem confirms that the best possible compositional AVaR that upper bounds  $\text{AVaR}_\varepsilon$  does involve compositions with the worst-case operator, in any node  $i$  that has probability at most  $1 - \varepsilon$ . Furthermore, it suggests, in a precise sense, that the compositional AVaR gets increasingly conservative as the risk measurement process proceeds in time: note that  $\gamma_i \geq \gamma_j, \forall j \in \mathcal{C}_i$ , and once node  $i$  requires a worst-case operator, so will any descendant of  $i$ , since  $\mathbb{P}(\mathcal{D}_i) \geq \mathbb{P}(\mathcal{D}_j), \forall j \in \mathcal{C}_i$ . In particular, *all* future stages are more conservative than the measurement at time  $t = 0$  (i.e., the root node), which exactly corresponds to the inconsistent evaluation  $\mu_I = \text{AVaR}_\varepsilon$ .

This last point may be of particular relevance when designing risk measures for use in dynamic financial settings: it suggests that regulators looking for safe counterparts (i.e., upper-bounds) for a static  $\text{AVaR}_\varepsilon$  should use risk measurement processes that are compositions of increasingly conservative  $\text{AVaR}_\varepsilon$  measurements.

**5. Conclusions.** In this paper, we examined two different paradigms for measuring risk in dynamic settings: a time-consistent formulation, whereby the risk assessments are designed so as to avoid naïve reversals of preferences in the measurement process, and a time-inconsistent one, which is easier to specify and calibrate from preference data. We discussed necessary and sufficient conditions under which one measurement uniformly bounds the other from above or below, and provided a notion of the multiplicative tightness with which one measure can be approximated by the other. We also showed that it is generally hard to compute the scaling factors even for distortion risk measures, but provided concrete examples when polynomial-time algorithms are possible.

**Appendix A: Submissives, Downward Monotone Closures and Anti-blocking Polyhedra.** In the current section, we discuss the important notion of the *down monotone closure* of a polytope, also known as its *anti-blocking polyhedron* or its *submissive*. Our exposition mostly follows Chapter 9 in Schrijver [52], to which we direct the interested reader for a more comprehensive treatment and references to related literature.

A polyhedron  $Q$  in  $\mathbb{R}^n$  is said to be *down-monotone* or of *anti-blocking type* if

$$Q \neq \emptyset, Q \subseteq \mathbb{R}_+^n, \text{ and } 0 \leq \mathbf{y} \leq \mathbf{x} \text{ and } \mathbf{x} \in Q \text{ imply } \mathbf{y} \in Q.$$

The following proposition summarizes a useful representation for down-monotone polyhedra.

**PROPOSITION 12.** *A polyhedron  $Q$  in  $\mathbb{R}^n$  is down-monotone if and only if there is a finite set  $\mathcal{I}$  of vectors  $\{\mathbf{a}_i\}_{i \in \mathcal{I}}$  and coefficients  $\{b_i\}_{i \in \mathcal{I}}$  such that  $\mathbf{a}_i \geq 0$ ,  $\mathbf{a}_i \neq 0$ ,  $b_i \geq 0$ ,  $\forall i \in \mathcal{I}$ , and*

$$Q = \{ \mathbf{x} \in \mathbb{R}^n : \mathbf{a}_i^T \mathbf{x} \leq b_i, \forall i \in \mathcal{I} \}.$$

*Proof.* The proof follows closely from the definitions. We omit it here, and direct the interested reader to [52].  $\square$

We remark that, whenever  $Q$  is full-dimensional, the right-hand sides  $b_i$  in the representation above can be taken to be strictly positive.

For any polyhedron  $Q \subseteq \mathbb{R}^n$ , we can define its *down-monotone closure*, also known as its *submissive*, by

$$\text{sub}(Q) \stackrel{\text{def}}{=} \{ \mathbf{y} \in \mathbb{R}_+^n : \exists \mathbf{x} \in Q, \mathbf{x} \geq \mathbf{y} \}. \quad (27)$$

It can be easily checked that  $\text{sub}(Q) = (Q + \mathbb{R}_-^n) \cap \mathbb{R}_+^n$ , and that  $\text{sub}(Q)$  is full-dimensional if and only if  $Q \setminus \{ \mathbf{x} \in \mathbb{R}^n : x_j = 0 \} \neq \emptyset$ , for all  $j \in [1, n]$  (see Balas and Fischetti [8]). A very interesting characterization of the down-monotone closure of a polyhedron is also possible in terms of the polar of the polyhedron  $P$ . However, since these results are not directly needed in our treatment here, we direct the interested reader to [8, 7] or Chapter 9 in [52] for more details.

Down-monotone polyhedra have been used for studying the strength of relaxations in integer programming and combinatorial optimization – see Goemans and Hall [29] are references therein. The following result is relevant for our purposes.

**THEOREM 11.** *Let  $P$  and  $Q$  be two nonempty, downward monotone polytopes in  $\mathbb{R}_+^n$ . Then*

1.  *$P \subseteq \alpha Q$  if and only if, for any nonnegative vector  $\mathbf{w} \in \mathbb{R}^n$ ,*

$$\max \{ \mathbf{w}^T \mathbf{x} : \mathbf{x} \in Q \} \geq \frac{1}{\alpha} \max \{ \mathbf{w}^T \mathbf{x} : \mathbf{x} \in P \}.$$

2. *If  $Q = \{ \mathbf{x} \in \mathbb{R}_+^n : \mathbf{a}_i^T \mathbf{x} \leq b_i, \forall i \in \mathcal{I} \}$ , where  $\mathbf{a}_i, b_i \geq 0$ , then*

$$\alpha^* = \max_{i \in \mathcal{I}} \frac{d_i}{b_i}, \text{ where } d_i \stackrel{\text{def}}{=} \max_{\mathbf{x} \in P} \mathbf{a}_i^T \mathbf{x}.$$

*Proof.* Part (1) is essentially Lemma 1 in [29]. Since the latter reference omits a proof, we include one below, for completeness. “ $\Rightarrow$ ” follows trivially. “ $\Leftarrow$ ” Note first that  $\alpha > 0$ . Assume (by contradiction) that  $\exists \bar{\mathbf{x}} \in P \setminus \alpha Q$ . Since  $Q$  is down-monotone, by Proposition 12, it can be written as  $Q = \{\mathbf{x} \in \mathbb{R}_+^n : \mathbf{a}_i^T \mathbf{x} \leq b_i, \forall i \in \mathcal{I}\}$ , where  $\mathbf{a}_i, b_i \geq 0, \forall i \in \mathcal{I}$ . Since  $\bar{\mathbf{x}} \notin \alpha Q$ , there exists  $j \in \mathcal{I}$  such that  $\mathbf{a}_j^T \bar{\mathbf{x}} > \alpha b_j$ . Since  $\bar{\mathbf{x}} \in P$ , we obtain the desired contradiction,  $\frac{1}{\alpha} \max\{\mathbf{a}_j^T \mathbf{x} : \mathbf{x} \in P\} \geq \frac{1}{\alpha} \mathbf{a}_j^T \bar{\mathbf{x}} > b_j \geq \max\{\mathbf{a}_j^T \mathbf{x} : \mathbf{x} \in Q\}$ .

Part (2) is exactly Theorem 2 in [29], to which we direct the reader for a complete proof.  $\square$

The above result shows that  $\alpha^*$  can be  $+\infty$ , which is the case if  $Q$  has a strictly smaller dimension than  $P$  (in this case, some  $b_i$  are 0, while the corresponding  $d_i$  are strictly positive [52]). However, if  $Q$  is full-dimensional,  $\alpha^*$  is always finite.

**Appendix B: Submodular Functions and Polymatroids** In this section of the Appendix, we discuss the basic properties of Choquet capacities in light of their connection with rank functions of polymatroids. The exposition is mainly based on volume B of [53] (Chapter 44) and Chapter 2 of [27] (Section 3.3), to which we direct the interested reader for more information.

Consider a ground set  $\Omega$  with  $|\Omega| = n$ , and let  $c$  be a set function on  $\Omega$ , that is,  $c : \mathcal{F} \mapsto \mathbb{R}$ , where  $\mathcal{F} = 2^\Omega$  is the set of all subsets of  $\Omega$ . The function  $c$  is called *submodular* if

$$c(T) + c(U) \geq c(T \cap U) + c(T \cup U), \forall T, U \in \mathcal{F}.$$

The function  $c$  is called *nondecreasing* if  $c(T) \leq c(U)$  whenever  $T \subseteq U \subseteq \Omega$ . For a given set function  $c$  on  $\Omega$ , we define the following two polyhedra

$$\begin{aligned} \mathcal{P}_c &\stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^{|\Omega|} : \mathbf{x} \geq 0, \mathbf{x}(S) \leq c(S), \forall S \subseteq \Omega\} \\ \mathcal{EP}_c &\stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^{|\Omega|} : \mathbf{x}(S) \leq c(S), \forall S \subseteq \Omega\}. \end{aligned} \quad (28)$$

Note that  $\mathcal{P}_c$  is nonempty if and only if  $c \geq 0$ , and that  $\mathcal{EP}_c$  is nonempty if and only if  $c(\emptyset) \geq 0$ . These conditions are trivially satisfied in our exposition, since all set functions  $c$  of interest are Choquet capacities, i.e., by Definition 1, they are nondecreasing and normalized,  $c(\emptyset) = 0, c(\Omega) = 1$ .

If  $c$  is a submodular function, then  $\mathcal{P}_c$  is called the *polymatroid associated with  $c$* , and  $\mathcal{EP}_c$  the *extended polymatroid associated with  $c$* . Note that a nonempty extended polymatroid is always unbounded, while a polymatroid is always a polytope, since  $0 \leq x_i \leq c(\{i\}), \forall i \in \Omega$ . The next theorem provides a very useful result concerning the set of tight constraints in the representation of  $\mathcal{EP}_c$ .

**THEOREM 44.2 IN [53].** *Let  $c$  be a submodular set function on  $\Omega$  and let  $\mathbf{x} \in \mathcal{EP}_c$ . Then the collection of sets  $U \subseteq \Omega$  satisfying  $\mathbf{x}(U) = c(U)$  is closed under taking unions and intersections.*

*Proof.* Suppose  $\mathbf{x}(T) = c(T)$  and  $\mathbf{x}(U) = c(U)$ . Then

$$c(T) + c(U) \geq c(T \cap U) + c(T \cup U) \geq \mathbf{x}(T \cap U) + \mathbf{x}(T \cup U) = \mathbf{x}(T) + \mathbf{x}(U) = c(T) + c(U),$$

hence equality must hold throughout, and  $\mathbf{x}(T \cap U) = c(T \cap U)$  and  $\mathbf{x}(T \cup U) = c(T \cup U)$ .  $\square$

A vector  $\mathbf{x} \in \mathcal{EP}_c$  (or in  $\mathcal{P}_c$ ) is called a *base vector* of  $\mathcal{EP}_c$  (or of  $\mathcal{P}_c$ ) if  $\mathbf{x}(\Omega) = c(\Omega)$ . The set of all base vectors is called the *base polytope* of  $c$  and is denoted by  $\mathcal{B}_c$ ,

$$\mathcal{B}_c \stackrel{\text{def}}{=} \{\mathbf{x} \in \mathbb{R}^{|\Omega|} : \mathbf{x}(S) \leq c(S), \forall S \subseteq \Omega, \mathbf{x}(\Omega) = c(\Omega)\}.$$

The following theorem summarizes several simple properties of  $\mathcal{B}_c$ , and its relation to  $\mathcal{EP}_c$  and  $\mathcal{P}_c$ .

**THEOREM 12.** *For any submodular function  $c$  satisfying  $c(\emptyset) = 0$ ,*

- (i)  $\mathcal{B}_c$  is a face of  $\mathcal{EP}_c$ , and is always a polytope.
- (ii)  $\mathcal{EP}_c = \mathcal{B}_c + \mathbb{R}_-^n$ , so that  $\mathcal{EP}_c$  and  $\mathcal{B}_c$  have the same extreme points.
- (iii)  $\mathcal{P}_c = \text{sub}(\mathcal{B}_c)$ .
- (iv) For any  $\lambda \geq 0$ ,  $\mathcal{B}_{\lambda c} = \lambda \cdot \mathcal{B}_c$ ,  $\mathcal{EP}_{\lambda c} = \lambda \cdot \mathcal{EP}_c$ , and  $\mathcal{P}_{\lambda c} = \lambda \cdot \mathcal{P}_c$ .

*Proof.* (i) The fact that  $\mathcal{B}_c$  is a face of  $\mathcal{EP}_c$  follows directly from the definitions. To see that  $\mathcal{B}_c$  is a polytope, note that, for any  $i \in \Omega$ ,  $x_i \leq c(\{i\})$ , and  $x_i = \mathbf{x}(\Omega) - \mathbf{x}(\Omega \setminus \{i\}) \geq c(\Omega) - c(\Omega \setminus \{i\})$ .

(ii) “ $\supseteq$ ” Follows trivially. “ $\subseteq$ ” Consider any  $\mathbf{y} \in \mathcal{EP}_c$ . Without loss of generality<sup>9</sup>, assume  $\mathbf{y}$  does not lie in the strict interior of  $\mathcal{EP}_c$ , and let  $\mathcal{I}_{\mathbf{y}} \stackrel{\text{def}}{=} \{S \in \mathcal{F} : \mathbf{y}(S) = c(S)\}$  denote the collection of sets corresponding to tight constraints at  $\mathbf{y}$ . If  $\Omega \in \mathcal{I}_{\mathbf{y}}$ , then  $\mathbf{y} \in \mathcal{B}_c$ , and the proof would be complete. Therefore, let us assume  $\Omega \notin \mathcal{I}_{\mathbf{y}}$ .

We claim that there exists  $s \in \Omega$  such that  $s \notin S, \forall S \in \mathcal{I}_{\mathbf{y}}$ . To see this, note that, if any  $s \in \Omega$  were contained in some  $S \in \mathcal{I}_{\mathbf{y}}$ , then  $\Omega \in \mathcal{I}_{\mathbf{y}}$ , since the set of tight constraints is closed under union and intersection, by Theorem 2. We can then consider the vector  $\mathbf{y}_{\lambda} = \mathbf{y} + \lambda \mathbf{1}_s$  for  $\lambda \geq 0$ . It is easy to test that, for small enough  $\lambda$ ,  $\mathbf{y}_{\lambda} \in \mathcal{EP}_c$ . By making  $\lambda$  sufficiently large, at least one constraint a set  $S$  containing  $s$  becomes tight, hence enlarging the set  $\mathcal{I}_{\mathbf{y}}$ . Repeating the argument for the point  $\mathbf{y}_{\lambda}$  recursively, we eventually recover a vector  $\tilde{\mathbf{y}}$  that belongs to  $\mathcal{B}_c$ . Since  $\tilde{\mathbf{y}} = \mathbf{y} + \boldsymbol{\xi}$  for some  $\boldsymbol{\xi} \geq 0$ , we have that  $\mathbf{y} \in \mathcal{B}_c + \mathbb{R}_-^n$ , which completes the proof of the first part of (ii). Since  $\mathbb{R}_-^n$  is a cone, and  $\mathcal{B}_c$  is a polytope, the representation exactly corresponds to the Motzkin decomposition of an arbitrary polyhedron, so that  $\text{ext}(\mathcal{EP}_c) = \text{ext}(\mathcal{B}_c)$ .

(iii) Follows immediately from (ii), since  $\mathcal{P}_c = \mathcal{EP}_c \cap \mathbb{R}_+^n = (\mathcal{B}_c + \mathbb{R}_-^n) \cap \mathbb{R}_+^n \stackrel{\text{def}}{=} \text{sub}(\mathcal{B}_c)$ .

(iv) Since  $\lambda c$  is also submodular, the results immediately follow from the definitions.  $\square$

A central result in the theory of submodularity, due to Edmonds, is that a linear function  $\mathbf{w}^T \mathbf{x}$  can be optimized over an (extended) polymatroid by an extension of the greedy algorithm. The following theorem summarizes the finding.

**THEOREM 13 (Theorem 44.3, Corollaries 44.3(a,b) in [53]).** *Let  $c : 2^{\Omega} \rightarrow \mathbb{R}$  be a submodular set function with  $c(\emptyset) = 0$ , and let  $\mathbf{w} \in \mathbb{R}_+^{|\Omega|}$ . Then the optimum solution of  $\max_{\mathbf{x} \in \mathcal{EP}_c} \mathbf{w}^T \mathbf{x}$  is*

$$\mathbf{x}(s_i) \stackrel{\text{def}}{=} c(\{s_1, \dots, s_i\}) - c(\{s_1, \dots, s_{i-1}\}), i \in [1, n],$$

where  $(s_1, \dots, s_n)$  is a permutation of the elements of  $\Omega$  such that  $\mathbf{w}(s_1) \geq \mathbf{w}(s_2) \geq \dots \geq \mathbf{w}(s_n)$ . If  $c$  is also nondecreasing, then the above  $\mathbf{x}$  is also an optimal solution to the problem  $\max_{\mathbf{x} \in \mathcal{P}_c} \mathbf{w}^T \mathbf{x}$ .

*Proof* The proof follows by duality arguments. We omit it here, and direct the interested reader to [53].  $\square$

In view of this result, the following characterization for the extreme points of  $\mathcal{B}_c$ ,  $\mathcal{EP}_c$  and  $\mathcal{P}_c$  is immediate.

**THEOREM 14.** *For a submodular set function  $c$  satisfying  $c(\emptyset) = 0$ , the extreme points of  $\mathcal{B}_c$  and  $\mathcal{EP}_c$  are given by*

$$x_{\sigma(i)} = c(\{\sigma(1), \dots, \sigma(i)\}) - c(\{\sigma(1), \dots, \sigma(i-1)\}), i \in [1, n],$$

where  $\sigma \in \Pi(\Omega)$  is any permutation of the elements of  $\Omega$ . When  $c$  is also nondecreasing, the extreme points of  $\mathcal{P}_c$  are given by

$$x_{\sigma(i)} = \begin{cases} c(\{\sigma(1), \dots, \sigma(i)\}) - c(\{\sigma(1), \dots, \sigma(i-1)\}) & \text{if } i \leq k, \\ 0 & \text{if } i > k, \end{cases}$$

where  $\sigma \in \Pi(\Omega)$  is any permutation of the elements of  $\Omega$ , and  $k$  ranges over  $[0, n]$ .

<sup>9</sup> Such a  $\mathbf{y}$  can always be obtained by adding a certain  $\boldsymbol{\xi} \geq 0$ , and if the resulting  $\mathbf{y} + \boldsymbol{\xi} \in \mathcal{B}_c + \mathbb{R}_-^n$ , then also  $\mathbf{y} \in \mathcal{B}_c + \mathbb{R}_-^n$ .

*Proof.* For a complete proof, we direct the reader to Theorem 3.22 in [27] and Section 44.6c in [53].

The previous result shows that there is a one-to-one correspondence between vertices of  $\mathcal{B}_c$  and permutations of  $[1, n]$ , and also that every inequality constraint in the characterization of  $\mathcal{B}_c$  is tight at some  $\mathbf{x} \in \mathcal{B}_c$ . The following corollary also immediately follows from the above result.

**COROLLARY 7.** *For any submodular  $c$  such that  $c(\emptyset) = 0$ ,  $\mathcal{B}_c \subset \mathbb{R}_+^n$  if and only if  $c$  is nondecreasing.*

*Proof.* “ $\Leftarrow$ ” Immediate, since  $\mathcal{B}_c$  is the convex hull of its extreme points, which (by Theorem 14) are nonnegative. “ $\Rightarrow$ ” Consider any two sets  $T \subset U \subseteq \Omega$ , and take a chain of sets  $S_1 \subset S_2 \subset \dots \subset S_{|U \setminus T|}$  such that  $S_1 = T$  and  $S_{|U \setminus T|} = U$ . By Theorem 14, there exists an extreme point  $\mathbf{x}$  of  $\mathcal{B}_c$  having elements  $c(S_{i+1}) - c(S_i)$ ,  $i \in [1, |U \setminus T| - 1]$  among some of its coordinates. Since  $\mathbf{x} \geq 0$ , we immediately obtain that  $c(U) - c(T) \geq 0$ .

**Appendix C: Technical Proofs.** This section contains several technical results from our analysis.

**PROPOSITION 6.** *Consider a (two-period) consistent, comonotonic risk measure  $\mu_C(Y) = \mu_1 \circ \mu_2$ , where  $\mu_t : \mathcal{X}_t \rightarrow \mathcal{X}_{t-1}$ . Then,*

1. *There exists  $\mathcal{Q}_C \subseteq \Delta^{|\Omega_2|}$  such that  $\mu_C(Y) \stackrel{\text{def}}{=} \max_{\mathbf{q} \in \mathcal{Q}_C} \mathbf{q}^T \mathbf{Y}$ ,  $\forall Y \in \mathcal{X}_2$ .*
2. *The set of measures  $\mathcal{Q}_C$  is given by*

$$\begin{aligned} \mathcal{Q}_C &\stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \exists \mathbf{p} \in \Delta^{|\Omega_1|}, \quad \begin{array}{l} \mathbf{p}(S) \leq c_1(S), \forall S \subseteq \Omega_1 \\ \mathbf{q}(U) \leq p_i \cdot c_{2|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1 \end{array} \right\} \\ &\equiv \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \exists \mathbf{p} \in \mathcal{B}_{c_1} : \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{B}_{p_i \cdot c_{2|i}}, \forall i \in \Omega_1 \right\}, \end{aligned}$$

where  $c_1 : 2^{|\Omega_1|} \rightarrow \mathbb{R}$  and  $c_{2|i} : 2^{|\mathcal{C}_i|} \rightarrow \mathbb{R}$ ,  $\forall i \in \Omega_1$  are Choquet capacities, and  $\mathcal{B}_{c_1}, \mathcal{B}_{c_{2|i}}$  are the base polytopes corresponding to  $c_1$  and  $c_{2|i}$ , respectively.

3. *The downward monotone closure of  $\mathcal{Q}_C$  is given by*

$$\begin{aligned} \text{sub}(\mathcal{Q}_C) &\stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \mathbb{R}_+^{|\Omega_2|} : \exists \mathbf{p} \in \mathbb{R}_+^{|\Omega_1|}, \quad \begin{array}{l} \mathbf{p}(S) \leq c_1(S), \forall S \subseteq \Omega_1, \\ \mathbf{q}(U) \leq p_i \cdot c_{2|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1 \end{array} \right\} \\ &= \left\{ \mathbf{q} \in \mathbb{R}_+^{|\Omega_2|} : \exists \mathbf{p} \in \mathcal{P}_{c_1} : \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{P}_{p_i \cdot c_{2|i}}, \forall i \in \Omega_1 \right\} \end{aligned}$$

where  $\mathcal{P}_{c_1}$  and  $\mathcal{P}_{p_i c_{2|i}}$  are the polymatroids associated with  $c_1$  and  $p_i c_{2|i}$ , respectively.

*Proof.* The first claim is a standard result in the literature [22, 6, 45], but we rederive it here together with the second claim, to keep the paper self-contained. To this end, recall that Definition 3 implies any DC comonotonic risk measure  $\mu_C$  can be written as  $\mu_1 \circ \mu_2$ , where  $\mu_1 : \mathcal{X}_1 \rightarrow \mathbb{R}$  is a first-period comonotonic risk measure, and  $\mu_2 \equiv (\mu^i)_{i \in \Omega_1}$ , where  $\mu^i : \mathbb{R}^{|\mathcal{C}_i|} \rightarrow \mathbb{R}$ ,  $\forall i \in \Omega_1$  are comonotonic risk measures. By the representation in Theorem 1, for any  $X_1 \in \mathcal{X}_1$  and  $X_2 \in \mathcal{X}_2$ , we have

$$\mu_1(X_1) = \max_{\mathbf{p} \in \mathcal{Q}_1} \mathbf{p}^T \mathbf{X}_1, \quad \mathcal{Q}_1 \stackrel{\text{def}}{=} \left\{ \mathbf{p} \in \Delta^{|\Omega_1|} : \mathbf{p}(S) \leq c_1(S), \forall S \subseteq \Omega_1 \right\}, \quad (29a)$$

$$\mu_2^i(X_2) = \max_{\mathbf{q} \in \mathcal{Q}_{2|i}} \mathbf{q}^T \mathbf{X}_2, \quad \mathcal{Q}_{2|i} \stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \mathbf{q}(U) \leq c_{2|i}(U), \forall U \subseteq \mathcal{C}_i; \mathbf{q}|_{\Omega_2 \setminus \mathcal{C}_i} = 0 \right\}. \quad (29b)$$

In particular,  $\mathcal{Q}_1 \equiv \mathcal{B}_{c_1}$ , and, similarly, the projection of the polytope  $\mathcal{Q}_{2|i}$  on the coordinates  $\mathcal{C}_i$  is exactly given by  $\mathcal{B}_{c_{2|i}}$ , for any  $i \in \Omega_1$ . From these relations, we have that  $\mu_C(Y) = \max_{\mathbf{q} \in \tilde{\mathcal{Q}}} \mathbf{q}^T \mathbf{Y}$ , where  $\tilde{\mathcal{Q}}$  has the following product form structure [55]:

$$\tilde{\mathcal{Q}} = \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \exists \mathbf{p} \in \mathcal{Q}_1, \exists \mathbf{q}^i \in \mathcal{Q}_{2|i}, \forall i \in \Omega_1, \text{ such that } \mathbf{q} = \sum_{i \in \Omega_1} p_i \mathbf{q}^i \right\}. \quad (30)$$

We now show that  $\tilde{\mathcal{Q}} = \mathcal{Q}_C$ , by double inclusion.

“ $\subseteq$ ” Consider any  $\mathbf{q} \in \tilde{\mathcal{Q}}$ , and let  $\mathbf{p} \in \mathcal{Q}_1$  and  $\mathbf{q}^i \in \mathcal{Q}_{2|i}$  denote the corresponding vectors in representation (30). Since  $\mathbf{q}_{\mathcal{C}_i} = p_i \cdot \mathbf{q}^i$ ,  $\forall i \in \Omega_1$ , and  $\mathbf{q}^i \in \mathcal{Q}_{2|i}$ ,  $\forall i \in \Omega_1$ , we trivially have that  $\mathbf{p}$  and  $\mathbf{q}$  satisfy the equations defining  $\mathcal{Q}_C$ .

“ $\supseteq$ ” Consider any  $\mathbf{q} \in \mathcal{Q}_C$ , and let  $\mathbf{p}$  be a corresponding measure satisfying the constraints for  $\mathcal{Q}_C$ . It can be readily checked that  $\mathbf{q}^i \stackrel{\text{def}}{=} \frac{\mathbf{q}_{\mathcal{C}_i}}{p_i} \in \mathcal{Q}_{2|i}$  (the only non-obvious constraint is  $\mathbf{1}^T \mathbf{q}^i = 1$ , which must hold, since, otherwise, we would have  $\sum_{i \in \Omega_1} \mathbf{q}(\mathcal{C}_i) = \sum_{i \in \Omega_1} p_i \mathbf{1}^T \mathbf{q}^i < \mathbf{p}(\Omega_1) = 1$ , contradicting  $\mathbf{q} \in \Delta^{|\Omega_2|}$ ). Therefore,  $\mathbf{q} = \sum_{i \in \Omega_1} p_i \mathbf{q}^i \in \tilde{\mathcal{Q}}$ . For completeness, we also note that  $\mathbf{q}^i = \frac{\mathbf{q}_{\mathcal{C}_i}}{p_i} \in \mathcal{Q}_{2|i} \Leftrightarrow \frac{\mathbf{q}_{\mathcal{C}_i}}{p_i} \in \mathcal{B}_{c_{2|i}} \Leftrightarrow \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{B}_{p_i \cdot c_{2|i}}$ , for any  $i \in \Omega_1$  (by part (iv) of Theorem 12 in Section B of the Appendix).

To prove the last claim, note that the two sets on the right being identical is immediate from the definition of the polymatroid associated with a rank function  $c$  (see Section B). As such, denote by  $\mathcal{A}$  the set on the right of the equation.

“ $\subseteq$ ”. Consider an arbitrary  $\mathbf{x} \in \text{sub}(\mathcal{Q}_C)$ . By definition,  $\mathbf{x} \geq 0$  and  $\exists \mathbf{q} \in \mathcal{Q}_C$  such that  $\mathbf{q} \geq \mathbf{x}$ . Let  $\mathbf{p}$  correspond to  $\mathbf{q}$  in the representation for  $\mathcal{Q}_C$ . To argue that  $\mathbf{x} \in \mathcal{A}$ , we show that the pair  $(\mathbf{p}, \mathbf{x})$  satisfies all the constraints defining  $\mathcal{A}$ . To this end, since  $\mathbf{p} \in \mathcal{B}_{c_1}$  (and  $\mathcal{B}_{c_1} \subset \mathbb{R}_+^{|\Omega_1|}$ ), we immediately have  $\mathbf{p} \in \mathcal{P}_{c_1}$ . Furthermore,  $\forall i \in \Omega_1$  and  $\forall U \subseteq \mathcal{C}_i$ , we have  $\mathbf{x}(U) \leq \mathbf{q}(U) \leq p_i \cdot c_{2|i}(U)$ , which proves that  $\mathbf{x} \in \mathcal{A}$ .

“ $\supseteq$ ”. Consider an arbitrary  $\mathbf{q} \in \mathcal{A}$ , and let  $\mathbf{p}$  be such that the pair  $(\mathbf{p}, \mathbf{q})$  satisfies all the constraints defining  $\mathcal{A}$ . Since  $\mathbf{p} \in \mathcal{P}_{c_1} \equiv \text{sub}(\mathcal{B}_{c_1})$ ,  $\exists \bar{\mathbf{p}} \in \mathcal{B}_{c_1}$  such that  $\bar{\mathbf{p}} \geq \mathbf{p} \geq 0$ . Furthermore,  $\mathbf{q}|_{\mathcal{C}_i} \in \mathcal{P}_{\bar{p}_i \cdot c_{2|i}} \equiv \text{sub}(\mathcal{B}_{\bar{p}_i \cdot c_{2|i}})$ , for any  $i \in \Omega_1$ . Therefore,  $\exists \bar{\mathbf{q}} \in \mathbb{R}_+^{|\Omega_2|}$  such that  $\bar{\mathbf{q}}|_{\mathcal{C}_i} \in \mathcal{B}_{\bar{p}_i \cdot c_{2|i}}$  and  $\bar{\mathbf{q}}|_{\mathcal{C}_i} \geq \mathbf{q}|_{\mathcal{C}_i} \geq 0$ , for any  $i \in \Omega_1$ . It can be readily checked that, by construction, the pair  $(\bar{\mathbf{p}}, \bar{\mathbf{q}})$  satisfies all the constraints defining  $\mathcal{Q}_C$ . Therefore, with  $\bar{\mathbf{q}} \in \mathcal{Q}_C$  and  $\bar{\mathbf{q}} \geq \mathbf{q} \geq 0$ , we must have  $\mathbf{q} \in \text{sub}(\mathcal{Q}_C)$ .  $\square$

**COROLLARY 4.** *For any pair of risk measures  $\mu_I$  and  $\mu_C$  as introduced in Section 3.2,*

1. *The inequality  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if*

$$\sum_{j=1}^{|\Omega_1|} \left[ c_1 \left( \cup_{k=1}^j s_k \right) - c_1 \left( \cup_{k=1}^{j-1} s_k \right) \right] \cdot c_{2|s_j}(U_{s_j}) \leq c \left( \cup_{i \in \Omega_1} U_i \right),$$

where  $(s_1, \dots, s_{|\Omega_1|})$  denotes any permutation of the elements of  $\Omega_1$ , and  $U_i \subseteq \mathcal{C}_i$  for any  $i \in \Omega_1$ .

2. *The inequality  $\mu_I(Y) \leq \mu_C(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if*

$$\begin{aligned} c(\cup_{i \in S} \mathcal{C}_i) &\leq c_1(S), \quad \forall S \subseteq \Omega_1, \\ \frac{c(U)}{c(U) + 1 - c(\Omega_2 \setminus \mathcal{C}_i \cup U)} &\leq c_{2,i}(U), \quad \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

*Proof.* The main idea proof behind the proof is to rewrite the results in Corollary 3 in terms of the extreme points of  $\mathcal{Q}_C$  and  $\mathcal{Q}_I$ , and then to suitably simplify the resulting problems.

To prove part (1), by Corollary 3, we have that  $\mu_C(Y) \leq \mu_I(Y)$ ,  $\forall Y \in \mathcal{X}_2$  holds if and only if

$$\max_{\mathbf{q} \in \text{ext}(\mathcal{Q}_C)} \mathbf{q}(S) \leq c(S), \quad \forall S \subseteq \Omega_2.$$

To this end, consider any  $S \subseteq \Omega_2$ , and partition it as  $S = \cup_{\ell \in \Omega_1} U_\ell$ , for some  $U_\ell \subseteq \mathcal{C}_\ell$ ,  $\forall \ell \in \Omega_1$ . The expression for  $\text{ext}(\mathcal{Q}_C)$  is given in Proposition 7, which we paste below for convenience

$$q_{\sigma_\ell(i)} = \left[ c_1 \left( \cup_{k=1}^\ell \pi(k) \right) - c_1 \left( \cup_{k=1}^{\ell-1} \pi(k) \right) \right] \cdot \left[ c_{2|\ell} \left( \cup_{k=1}^i \sigma_\ell(k) \right) - c_{2|\ell} \left( \cup_{k=1}^{i-1} \sigma_\ell(k) \right) \right], \quad \forall i \in [1, |\mathcal{C}_\ell|], \forall \ell \in \Omega_1.$$

where  $\pi$  is any permutation of  $\Omega_1$ , and  $\sigma_\ell$  is any permutation of  $\mathcal{C}_\ell$ , for each  $\ell \in \Omega_1$ .

Consider a fixed permutation  $\pi \in \Omega_1$ . We claim that the permutation  $\sigma_\ell$  yielding a maximal value of  $\mathbf{q}(U_\ell)$  is always of the form  $(\sigma(U_\ell), \sigma(\mathcal{C}_\ell \setminus U_\ell))$ , i.e., it has the elements of  $U_\ell$  in the first  $|U_\ell|$

positions. This is because the functions  $c_{2|\ell}$  are submodular, so that  $c(U_\ell) - c(\emptyset) \geq c(U_\ell \cup A) - c(A)$ , for any  $A \subseteq \mathcal{C}_\ell \setminus U_\ell$ . With this recognition, the optimal permutations  $\sigma_\ell$  always result in  $\mathbf{q}(U_\ell) = \left[ c_1(\cup_{k=1}^\ell \pi(k)) - c_1(\cup_{k=1}^{\ell-1} \pi(k)) \right] c_{2|\ell}(U_\ell)$ . Maximizing over all permutations  $\pi \in \Pi(\Omega_1)$  then leads to the first set of desired conditions.

To prove part (2), one can use the expression from Corollary 3, and show that it reduces to the desired condition. Instead, we find it more convenient to work with the results of Lemma 2 concerning  $\widehat{\mu}_I$ , the tightest possible coherent upper bound to  $\mu_I$ . To this end, first recall the representation for  $\mathcal{Q}_C$  in Proposition 6, pasted below for convenience:

$$\mathcal{Q}_C = \left\{ \mathbf{q} \in \Delta^{|\Omega_2|} : \exists \mathbf{p} \in \mathcal{B}_{c_1} : \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{B}_{p_i \cdot c_{2|i}}, \forall i \in \Omega_1 \right\}.$$

Lemma 2 implies that  $\mu_I(Y) \leq \mu_C(Y) \Leftrightarrow \widehat{\mathcal{Q}}_{\mu_I}^i \subseteq \mathcal{B}_{c_{t|i}}, \forall i \in \Omega_{t-1}, \forall t \in [1, 2]$ . Here,  $\widehat{\mathcal{Q}}_{\mu_I}^i$  are the one-step conditional risk measures yielding  $\widehat{\mu}_I$ , and are given by (10). This is equivalent to

$$\max_{\mathbf{q} \in \mathcal{Q}_I} \mathbf{q}(\cup_{i \in S} \mathcal{C}_i) \leq c_1(S), \forall S \subseteq \Omega_1 \quad (*)$$

$$\max_{\mathbf{q} \in \mathcal{Q}_I : \mathbf{q}(\mathcal{C}_i) \neq 0} \frac{\mathbf{q}(U)}{\mathbf{q}(\mathcal{C}_i)} \leq c_{2|i}(U), \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1 \quad (**)$$

We now argue that (\*) and (\*\*) are equivalent to the conditions in part (2). Recalling the description of  $\mathcal{Q}_I$  in Proposition 5, and the fact that any inequality  $\mathbf{q}(S) \leq c(S)$  is tight at some set  $S$  (also see Theorem 14), it can be seen that the maximum value of  $\mathbf{q}(\cup_{i \in S} \mathcal{C}_i)$  in (\*) is exactly  $c(\cup_{i \in S} \mathcal{C}_i)$ , which yields the first desired condition. The proof that (\*\*) are equivalent to the second condition is the subject of Proposition 13 below, which completes our proof.  $\square$

PROPOSITION 13. *Consider any  $i \in \Omega_{t-1}$  for some  $t \in [1, T]$ . Then, for any  $U \subseteq \mathcal{C}_i$ , we have*

$$\max_{\mathbf{q} \in \mathcal{Q}_I : \mathbf{q}(\mathcal{C}_i) \neq 0} \frac{\mathbf{q}(\mathcal{D}_U)}{\mathbf{q}(\mathcal{D}_i)} = \frac{c(\mathcal{D}_U)}{c(\mathcal{D}_U) + 1 - c(\Omega_T \setminus \mathcal{D}_i \cup \mathcal{D}_U)}. \quad (31)$$

*Proof.* Since the problem on the left is a fractional linear program, the maximum is reached at an extreme point of  $\mathcal{Q}_I$  [13]. Note also that the objective is increasing in any  $q_j, j \in \mathcal{D}_U$ , and decreasing in any  $q_s, s \in \mathcal{D}_i \setminus \mathcal{D}_U$ .

Recalling the expression for  $\text{ext}(\mathcal{Q}_I)$  in Proposition 7,

$$q_{\sigma(i)} = c(\cup_{k=1}^i \sigma(k)) - c(\cup_{k=1}^{i-1} \sigma(k)), \forall i \in [1, |\Omega_T|],$$

let  $\mathbf{v}^\sigma \in \text{ext}(\mathcal{Q}_I)$  be the extreme point corresponding to  $\sigma \in \Pi(\Omega_T)$ . We claim that there exists an optimal solution in (31) such that the permutation  $\sigma$  is of the form

$$\begin{aligned} \mathcal{D}_U &= \{\sigma(1), \dots, \sigma(|\mathcal{D}_U|)\} \\ \mathcal{D}_i \setminus \mathcal{D}_U &= \{\sigma(|\Omega_T \setminus \mathcal{D}_i \cup \mathcal{D}_U| + 1), \dots, \sigma(|\Omega_T|)\}, \end{aligned} \quad (32)$$

i.e., the elements of  $\mathcal{D}_U$  appear in the first  $|\mathcal{D}_U|$  positions of  $\sigma$ , and the elements of  $\mathcal{D}_i \setminus \mathcal{D}_U$  appear in the last positions of  $\sigma$ .

The proof involves a repeated interchange argument. We first argue that there exists an optimal permutation  $\sigma$  such that the elements of  $\mathcal{D}_U$  appear *before* those of  $\mathcal{D}_i \setminus \mathcal{D}_U$ .

To see this, consider any permutation  $\sigma$  such that  $\mathbf{v}^\sigma$  is optimal in (31), yet there exist  $j \in \mathcal{D}_U$  and  $s \in \mathcal{D}_i \setminus \mathcal{D}_U$  such that  $j = \sigma(k)$ ,  $s = \sigma(\bar{k})$ , and  $\bar{k} < k$ . In fact, let  $k$  be the smallest, and  $\bar{k}$  the largest such index among all indices satisfying the property (this ensures that there are no indices from  $\mathcal{D}_i$  appearing in  $\sigma$  between  $\bar{k}$  and  $k$ ). Consider a new permutation  $\pi$  where the positions  $k$  and  $\bar{k}$  are interchanged, and let  $\mathbf{v}^\pi$  denote the corresponding vertex of  $\mathcal{Q}_I$ . By submodularity of  $c$ ,

$$v_j^\pi \stackrel{\text{def}}{=} c(\cup_{\ell=1}^{\bar{k}} \sigma(\ell)) - c(\cup_{\ell=1}^{\bar{k}-1} \sigma(\ell)) \geq c(\cup_{\ell=1}^k \sigma(\ell)) - c(\cup_{\ell=1}^{k-1} \sigma(\ell)) \stackrel{\text{def}}{=} v_j^\sigma.$$



By a similar argument,  $v_s^\pi \leq v_s^\sigma$ . Furthermore, by construction,  $v_r^\pi = v_r^\sigma, \forall r \in \mathcal{D}_i \setminus \{j, s\}$ , since no indices from  $\mathcal{D}_i$  appear between  $\bar{k}$  and  $k$ . Therefore, we have  $\mathbf{v}^\pi(\mathcal{D}_U) \geq \mathbf{v}^\sigma(\mathcal{D}_U)$ , and  $\mathbf{v}^\pi(\mathcal{D}_i \setminus \mathcal{D}_U) \leq \mathbf{v}^\sigma(\mathcal{D}_i \setminus \mathcal{D}_U)$ , so that the objective at  $\mathbf{v}^\pi$  is at least as large as at  $\mathbf{v}^\sigma$ . Repeating the argument as often as needed, we obtain an optimal permutation satisfying the desired property.

Having argued that (w.l.o.g.)  $\sigma$  contains the elements of  $\mathcal{D}_U$  before those of  $\mathcal{D}_i \setminus \mathcal{D}_U$ , a similar interchange argument can be done with respect to  $\Omega_T \setminus \mathcal{D}_i$ , to reach the conclusion (32). The final result of the lemma exactly denotes the value corresponding to such a configuration (it follows immediately by recognizing the telescoping sums appearing in the expressions).  $\square$

**PROPOSITION 9.** *Consider any distortion risk measure  $\mu_I : \mathcal{X}_2 \rightarrow \mathbb{R}$ , and the time-consistent, comonotonic measures  $\mu_I \circ \mathbb{E}$  and  $\mathbb{E} \circ \mu_I$ , where  $\mathbb{E}$  denotes the conditional expectation operator. Then, for any cost  $Y \in \mathcal{X}_2$ ,*

$$(\mu_I \circ \mathbb{E})(Y) \leq \mu_I(Y) \quad \text{and} \quad (\mathbb{E} \circ \mu_I)(Y) \leq \mu_I(Y).$$

*Proof.* First note that both measures are readily time-consistent and comonotonic, by Definition 3. The proof entails arguing that these choices correspond to Choquet capacities that verify the first set of conditions in Corollary 4.

To this end, let  $\Psi$  denote the distortion function corresponding to the (distortion) measure  $\mu_I$ , i.e., the Choquet capacity is given by  $c(S) = \Psi(\mathbb{P}(S)), \forall S \subseteq \Omega_2$ , where  $\Psi : [0, 1] \rightarrow [0, 1]$  is concave, nondecreasing, with  $\Psi(0) = 0, \Psi(1) = 1$ . Recall from Section 4.2 that the (compositional) risk measure  $\mu_I \circ \mathbb{E}$  (or, more correctly,  $\mu_I^\dagger \circ \mathbb{E}$ ) exactly corresponds to the following choice of Choquet capacities for the first and second stage, respectively:

$$\begin{aligned} c_1 : 2^{\Omega_1} &\rightarrow \mathbb{R}, c_1(S) = \Psi\left(\sum_{i \in S} \mathbb{P}(\mathcal{C}_i)\right), \forall S \subseteq \Omega_1 \\ c_{2|i} : 2^{\mathcal{C}_i} &\rightarrow \mathbb{R}, c_{2|i}(U_i) = \frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}, \forall U_i \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

Note that the same distortion function  $\Psi$  (yielding the risk measure  $\mu_I$ ) is applied in the first stage, but to the appropriate conditional probability measure. The second stage is simply a standard conditional expectation.

With  $p_i \stackrel{\text{def}}{=} \mathbb{P}(\mathcal{C}_i)$  and  $u_i \stackrel{\text{def}}{=} \mathbb{P}(U_i)$ , the desired condition in Corollary 4 becomes:

$$\sum_{i=1}^{|\Omega_1|} \frac{\Psi\left(\sum_{j=1}^i p_{\sigma(j)}\right) - \Psi\left(\sum_{j=1}^{i-1} p_{\sigma(j)}\right)}{p_{\sigma(i)}} u_{\sigma(i)} \leq \Psi\left(\sum_{i=1}^{|\Omega_1|} u_{\sigma(i)}\right), \forall \sigma \in \Pi(\Omega_1), \forall u_i \in [0, p_i], \forall i \in \Omega_1. (*)$$

To see this, one can use the decreasing marginal returns property of  $\Psi$ , i.e.,

$$\frac{\Psi(y_2) - \Psi(y_1)}{y_2 - y_1} \leq \frac{\Psi(x_2) - \Psi(x_1)}{x_2 - x_1}, \forall x_1 < x_2, x_1 \leq y_1, x_2 \leq y_2, y_1 < y_2,$$

to argue that  $\frac{\Psi\left(\sum_{j=1}^i p_{\sigma(j)}\right) - \Psi\left(\sum_{j=1}^{i-1} p_{\sigma(j)}\right)}{p_{\sigma(i)}} \leq \frac{\Psi\left(\sum_{j=1}^i u_{\sigma(j)}\right) - \Psi\left(\sum_{j=1}^{i-1} u_{\sigma(j)}\right)}{u_{\sigma(i)}}$ . Replacing this in the left-hand side of (\*) and telescoping the sum directly yields the desired result.

In a similar fashion, the risk measure  $\mathbb{E} \circ \mu_I$  corresponds to a choice of capacities

$$\begin{aligned} c_1 : 2^{\Omega_1} &\rightarrow \mathbb{R}, c_1(S) = \sum_{i \in S} \mathbb{P}(\mathcal{C}_i), \forall S \subseteq \Omega_1 \\ c_{2|i} : 2^{\mathcal{C}_i} &\rightarrow \mathbb{R}, c_{2|i}(U_i) = \Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right), \forall U_i \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

With the same notation  $p_i \stackrel{\text{def}}{=} \mathbb{P}(\mathcal{C}_i)$  and  $u_i \stackrel{\text{def}}{=} \mathbb{P}(U_i)$ , the conditions to test become:

$$\sum_{i=1}^{|\Omega_1|} p_{\sigma(i)} \Psi\left(\frac{u_{\sigma(i)}}{p_{\sigma(i)}}\right) \leq \Psi\left(\sum_{i=1}^{|\Omega_1|} u_i\right), \forall \sigma \in \Pi(\Omega_1), \forall u_i \in [0, p_i], \forall i \in \Omega_1.$$

These are readily true, since  $\{p_i\}_{i \in \Omega_1}$  are convex combination coefficients, and  $\Psi$  is concave.  $\square$

**THEOREM 7.** *Consider a uniform scenario tree, i.e.,  $|\Omega_1| = N$ ,  $|\mathcal{C}_i| = N$ ,  $\forall i \in \Omega_1$ , under a uniform reference measure. Then, for any distortion risk measure  $\mu_I$ , we have*

$$\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^* = \alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* = N \cdot \max\left\{\frac{\Psi(1/N^2)}{\Psi(1/N)}, \frac{\Psi(2/N^2)}{\Psi(2/N)}, \dots, \Psi(1/N)\right\}.$$

Before presenting the proof, we introduce two lemmas that outline several relevant properties for the two expressions that need to be compared. To fix ideas, assume the distortion risk measure  $\mu_I$  is given by a concave distortion function  $\Psi : [0, 1] \rightarrow [0, 1]$ . To this end, by applying the result in Theorem 4, our goal is to argue that

$$\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* \stackrel{\text{def}}{=} \max_{\mathbf{q} \in \text{ext}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \frac{\sum_{i \in S} \max_{U_i \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U_i)}{\Psi(\mathbb{P}(U_i)/\mathbb{P}(\mathcal{C}_i))}}{\sum_{i \in S} \mathbb{P}(\mathcal{C}_i)} = \max_{\mathbf{q} \in \text{ext}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \frac{\sum_{i \in S} \max_{U_i \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U_i)}{\mathbb{P}(U_i)/\mathbb{P}(\mathcal{C}_i)}}{\Psi(\sum_{i \in S} \mathbb{P}(\mathcal{C}_i))} \stackrel{\text{def}}{=} \alpha_{\mu_I \circ \mathbb{E}, \mu_I}^*. \quad (33)$$

The following lemma discusses the factor  $\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^*$  in the expression above.

**LEMMA 4.** *Consider the maximization problems yielding  $\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^*$  in (33). We claim that:*

1. *For any given  $\mathbf{q} \in \mathcal{Q}_I$ , the inner maximization over  $S \subseteq \Omega_1$  is reached at a singleton set  $S = \{i\}$  for some  $i \in \Omega_1$ .*
2. *The optimal  $\mathbf{q} \in \mathcal{Q}_I$  in the outer maximization always corresponds to a permutation  $\sigma \in \Pi(\Omega_2)$  satisfying the property*

$$\{\sigma(1), \dots, \sigma(N)\} = \mathcal{C}_i, \quad (34)$$

for some  $i \in \Omega_1$ . That is, the first  $N$  elements in the permutation belong to the same subtree  $\mathcal{C}_i$ .

3. *For any fixed  $i \in \Omega_1$ ,*

$$\max_{\mathbf{q} \in \text{ext}(\mathcal{Q}_I)} \max_{U_i \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U_i)}{\Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right)} = \max_{U_i \subseteq \mathcal{C}_i} \frac{\Psi(\mathbb{P}(U_i))}{\Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right)}.$$

4.  $\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* = \max_{i \in \Omega_1} \max_{U_i \subseteq \mathcal{C}_i} \frac{\Psi(\mathbb{P}(U_i))}{\mathbb{P}(\mathcal{C}_i) \Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right)}.$

*Proof of Lemma 4.* Claim (1) follows from the mediant inequality. To see this, for a fixed  $\mathbf{q}$ , let  $v_i \stackrel{\text{def}}{=} \max_{U_i \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U_i)}{\Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right)}$ ,  $\forall i \in \Omega_1$ , and note that the maximum over  $S \subseteq \Omega_1$  is achieved at any singleton  $\{i\} \subseteq \arg \max\{v_\ell / \mathbb{P}(\mathcal{C}_\ell) : \ell \in \Omega_1\}$ .

To see Claim (2), first recall that the set  $\text{ext}(\mathcal{Q}_I)$  corresponds to all possible permutations of  $\mathcal{C}_2$  (Proposition 7). By Claim (1), since the inner maximum always occurs at a singleton  $i^*(\mathbf{q})$ , the optimal  $\mathbf{q}^*$  must be such that components in  $\mathcal{C}_{i^*(\mathbf{q}^*)}$  are “as large as possible”. Due to the concavity of  $\Psi$ , this occurs when they appear in the first  $N$  positions in the permutation  $\sigma$  (also see the proof of Corollary 4).

Claim (3) follows directly from Claim (2), by switching the order of the two maximizations, and using the expression for the extreme points of  $\mathcal{Q}_I$  from Proposition 7.

Claim (4) follows from Claims (1) and (3), after switching the order of the maximizations over  $S$  and  $\mathbf{q}$ .  $\square$

The following lemma similarly summarizes properties of the second quantity of interest,  $\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^*$ .

LEMMA 5. *Consider the maximization problems yielding  $\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^*$  in (33). We claim that:*

1. *For any given  $\mathbf{q} \in \mathcal{Q}_I$ , and any  $i \in \Omega_1$ , the inner maximization over  $U_i \subseteq \mathcal{C}_i$  is reached at a singleton set  $U_i = \{j\}$  for some  $j \in \mathcal{C}_i$ .*

2. *Fix  $S \subseteq \Omega_1$ . The optimal  $\mathbf{q}^*(S) \in \mathcal{Q}_I$  corresponds to a permutation  $\sigma^S \in \Pi(\Omega_2)$  such that*

$$\nexists j_{1,2} \in \{1, \dots, |S|\} \text{ such that } \sigma(j_1), \sigma(j_2) \in \mathcal{C}_i, \text{ for some } i \in \Omega_1.$$

*In other words, the first  $|S|$  elements in the permutation  $\sigma$  belong to distinct subtrees  $\mathcal{C}_i$ .*

3. *Under the same setup as (2), the first  $|S|$  elements in  $\sigma^S \in \Pi(\Omega_2)$  correspond to the minimum-probability in their respective subtree, i.e.,*

$$\forall k \in [1, |S|], \quad \sigma^S(k) \in \arg \min_{j \in \mathcal{C}_i} \mathbb{P}_j, \text{ where } i \text{ is such that } \sigma(k) \in \mathcal{C}_i.$$

4. *Let  $m(i) \stackrel{\text{def}}{=} \arg \min_{j \in \mathcal{C}_i} \mathbb{P}_j$ . Then*

$$\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^* = \max_{S \subseteq \Omega_1} \max_{\sigma \in \Pi(S)} \frac{\sum_{i=1}^{|S|} \mathbb{P}(\mathcal{C}_{m(\sigma(i))}) \frac{\Psi\left(\sum_{k=1}^i \mathbb{P}_{m(\sigma(k))}\right) - \Psi\left(\sum_{k=1}^{i-1} \mathbb{P}_{m(\sigma(k))}\right)}{\mathbb{P}_{m(\sigma(i))}}}{\Psi\left(\sum_{i=1}^{|S|} \mathbb{P}(\mathcal{C}_{m(\sigma(i))})\right)}.$$

*Proof.* Claim (1) follows, again, by the mediant inequality. The logic is the same as in Claim (1) of Lemma 4, and is omitted.

Claim (2) follows from Claim (1), and by recognizing again that  $\mathbf{q}$  should have components “as large as possible” in the singletons  $j$  that yield the maximums.

To see Claim (3), first note that Claim (2) allows restricting attention to permutations  $\sigma^S$  that have elements from distinct subtrees in the first  $|S|$  components. For any such  $\sigma(j)$ , with  $j \in \{1, \dots, |S|\}$ ,

$$q_{\sigma(j)} = \frac{\Psi(\mathbb{P}_{\sigma(j)} + \sum_{k=1}^{j-1} \mathbb{P}_{\sigma(k)}) - \Psi(\sum_{k=1}^{j-1} \mathbb{P}_{\sigma(k)})}{\mathbb{P}_{\sigma(j)}/\mathbb{P}(\mathcal{C}_i)},$$

where  $\sigma(j) \in \mathcal{C}_i$ . By the concavity of  $\Psi$ , the above expression is decreasing in  $\mathbb{P}_{\sigma(j)}$ , which implies that  $\sigma(j)$  always corresponds to the element in  $\mathcal{C}_i$  with smallest probability.

Claim (4) follows from the previous three.  $\square$

With the previous results, we are now ready to provide a complete proof for our desired result, namely that under a uniform reference measure,  $\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* = \alpha_{\mu_I \circ \mathbb{E}, \mu_I}^*$ .

*Proof of Theorem 7.* By Lemma 4,  $\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* = \max_{i \in \Omega_1} \max_{U_i \subseteq \mathcal{C}_i} \frac{\Psi(\mathbb{P}(U_i))}{\mathbb{P}(\mathcal{C}_i) \Psi\left(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\right)}$ . For a uniform reference measure, due to the symmetry, this expression becomes

$$\alpha_{\mathbb{E} \circ \mu_I, \mu_I}^* = N \cdot \max \left\{ \frac{\Psi(1/N^2)}{\Psi(1/N)}, \frac{\Psi(2/N^2)}{\Psi(2/N)}, \dots, \Psi(1/N) \right\}.$$

Similarly, by Lemma 5,  $\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^* = \max_{S \subseteq \Omega_1} \max_{\sigma \in \Pi(S)} \frac{\sum_{i=1}^{|S|} \mathbb{P}(\mathcal{C}_{m(\sigma(i))}) \frac{\Psi\left(\sum_{k=1}^i \mathbb{P}_{m(\sigma(k))}\right) - \Psi\left(\sum_{k=1}^{i-1} \mathbb{P}_{m(\sigma(k))}\right)}{\mathbb{P}_{m(\sigma(i))}}}{\Psi\left(\sum_{i=1}^{|S|} \mathbb{P}(\mathcal{C}_{m(\sigma(i))})\right)}$ ,

which becomes, under uniform reference measure,

$$\alpha_{\mu_I \circ \mathbb{E}, \mu_I}^* = N \cdot \max \left\{ \frac{\Psi(1/N^2)}{\Psi(1/N)}, \frac{\Psi(2/N^2)}{\Psi(2/N)}, \dots, \Psi(1/N) \right\}.$$

Comparing the two expressions above immediately yields the desired equality.  $\square$

**PROPOSITION 10.** Consider any distortion risk measure  $\mu_I$ , and the time-consistent, comonotonic measures  $\mu_I \circ \max$  and  $\max \circ \mu_I$ , where  $\max$  denotes the conditional worst-case operator. Then:

(i) For any cost  $Y \in \mathcal{X}_2$ ,  $\mu_I(Y) \leq (\mu_I \circ \max)(Y)$ .

(ii) There exists a choice of  $\mu_I$  and of random costs  $Y_{1,2} \in \mathcal{X}_2$  such that  $(\max \circ \mu_I)(Y_1) < \mu_I(Y_1)$  and  $(\max \circ \mu_I)(Y_2) > \mu_I(Y_2)$ .

*Proof.* Let the Choquet capacity yielding the distortion measure  $\mu_I$  be of the form  $c(S) = \Psi(\mathbb{P}(S))$ ,  $\forall S \subseteq \Omega_2$ . We show Part (i) of the corollary by checking the conditions of Corollary 4. Recall that the risk measure  $\mu_I^1 \circ \max$  corresponds to a choice of capacities

$$\begin{aligned} c_1 : 2^{\Omega_1} &\rightarrow \mathbb{R}, c_1(S) = \Psi\left(\sum_{i \in S} \mathbb{P}(\mathcal{C}_i)\right), \forall S \subseteq \Omega_1 \\ c_{2|i} : 2^{\mathcal{C}_i} &\rightarrow \mathbb{R}, c_{2|i}(U_i) = 1, \forall U_i \neq \emptyset \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

The conditions to check from Corollary 4 are

$$\begin{aligned} \Psi(\mathbb{P}(\cup_{i \in S} \mathcal{C}_i)) &\leq \Psi\left(\sum_{i \in S} \mathbb{P}(\mathcal{C}_i)\right), \quad \forall S \subseteq \Omega_1, \\ \frac{\Psi(\mathbb{P}(U))}{\Psi(\mathbb{P}(U)) + 1 - \Psi(\mathbb{P}(\Omega_2 \setminus \mathcal{C}_i \cup U))} &\leq 1, \quad \forall U \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

The first inequality holds since  $\mathbb{P}(\cup_{i \in S} \mathcal{C}_i) = \sum_{i \in S} \mathbb{P}(\mathcal{C}_i)$ . The second inequality readily follows since  $\Psi$  is upper bounded by 1.  $\square$

**PROPOSITION 11.** Consider a uniform scenario tree, i.e.,  $|\Omega_1| = N$ ,  $|\mathcal{C}_i| = N$ ,  $\forall i \in \Omega_1$ , under a uniform reference measure. Then, for any distortion risk measure  $\mu_I$ , we have

$$\alpha_{\mu_I, \mu_I \circ \max}^* = \max\left\{\frac{\Psi(1/N)}{\Psi(1/N^2)}, \frac{\Psi(2/N)}{\Psi(2/N^2)}, \dots, \frac{1}{\Psi(1/N)}\right\}.$$

*Proof.* Recall that the risk measure  $\mu_C \equiv \mu_I \circ \max$  (or, more correctly,  $\mu_I^1 \circ \max$ ) corresponds to a choice of capacities

$$\begin{aligned} c_1 : 2^{\Omega_1} &\rightarrow \mathbb{R}, c_1(S) = \Psi\left(\sum_{i \in S} \mathbb{P}(\mathcal{C}_i)\right) \equiv \Psi\left(\frac{|S|}{N}\right), \forall S \subseteq \Omega_1 \\ c_{2|i} : 2^{\mathcal{C}_i} &\rightarrow \mathbb{R}, c_{2|i}(U_i) = 1, \forall U_i \neq \emptyset \subseteq \mathcal{C}_i, \forall i \in \Omega_1. \end{aligned}$$

By Theorem 4, the optimal scaling factor is given by  $\alpha_{\mu_I, \mu_C}^* = \max_{\mathbf{q} \in \mathcal{Q}_C} \max_{S \subseteq \Omega_2} \mathbf{q}(S) / \Psi\left(\frac{|S|}{N^2}\right)$ . Let us switch the order of the maximizations, and fix an arbitrary  $S = \cup_{i \in \Omega_1} U_i \subseteq \Omega_2$ . Using the representation of  $\mathcal{Q}_C$  provided by Proposition 6, it can be readily seen that  $\mathbf{q}(U_i) = 0$  if  $U_i = \emptyset$ , and  $\mathbf{q}(U_i) \leq p_i$ , otherwise, where  $\mathbf{p}(S) \leq c_1(S)$ ,  $\forall S \subseteq \Omega_1$ . Therefore,

$$\max_{\mathbf{q} \in \mathcal{Q}_C} \mathbf{q}(S) = c_1(S) = \Psi\left(\frac{|S|}{N}\right),$$

which, when used in the expression for  $\alpha_{\mu_I, \mu_C}^*$ , immediately leads to the desired result.

**Acknowledgments.** The authors are grateful to Viswanath Nagarajan and Nikhil Bansal for their help with the hardness proof in Section 4.1.

## References

- [1] Acciaio, Beatrice, Irina Penner. 2011. Dynamic Risk Measures. Giulia Di Nunno, Bernt Øksendal, eds., *Advanced Mathematical Methods for Finance*, chap. 1. Springer, 1–34.
- [2] Acerbi, Carlo. 2002. Spectral Measures of Risk: A coherent Representation of Subjective Risk Aversion. *Journal of Banking & Finance* **26**(7) 1505 – 1518.
- [3] Acerbi, Carlo. 2004. Coherent Representations of Subjective Risk Aversion. Giorgio Szegő, ed., *Risk Measures for the 21st Century*, chap. 13. John Wiley & Sons, 147.
- [4] Ahmed, Shabbir, Ulaş Çakmak, Alexander Shapiro. 2007. Coherent risk measures in inventory problems. *European Journal of Operational Research* **182** 226–238.
- [5] Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, David Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* **9**(3) 203–228.
- [6] Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, David Heath, Hyejin Ku. 2007. Coherent multi-period risk adjusted values and bellman’s principle. *Annals of Operations Research* **152** 5–22.
- [7] Balas, Egon, Alexander Bockmayr, Nicolai Pizaruk, Laurence Wolsey. 2004. On unions and dominants of polytopes. *Mathematical Programming* **99** 223–239.
- [8] Balas, Egon, Matteo Fischetti. 1996. On the monotoneization of polyhedra. *Mathematical Programming* **78** 59–84.
- [9] Basak, S, A Shapiro. 2001. Value-at-risk-based risk management: optimal policies and asset prices. *Review of Financial Studies* **14**(2) 371–405.
- [10] Basak, Suleyman, Georgy Chabakauri. 2010. Dynamic mean-variance asset allocation. *Review of Financial Studies* **23**(8) 2970–3016.
- [11] Ben-Tal, Aharon, Marc Teboulle. 2007. An old-new concept of convex risk measures: The optimized certainty equivalent. *Mathematical Finance* **17**(3) 449–476.
- [12] Bertsimas, Dimitris, David B. Brown. 2009. Constructing uncertainty sets for robust linear optimization. *Operations Research* **57**(6) 1483–1495.
- [13] Boyd, Stephen, Lieven Vandenbergh. 2004. *Convex Optimization*. Cambridge University Press.
- [14] Brazauskas, Vytautas, Bruce L. Jones, Madan L. Puri, Ričardas Zitikis. 2008. Estimating conditional tail expectation with actuarial applications in view. *Journal of Statistical Planning and Inference* **138**(11) 3590 – 3604.
- [15] Cheridito, Patrick, Freddy Delbaen, Michael Kupper. 2006. Dynamic monetary risk measures for bounded discrete-time processes. *Electronic Journal of Probability* **11** 57–106.
- [16] Cheridito, Patrick, Michael Kupper. 2011. Composition of time-consistent dynamic monetary risk measures in discrete time. *International Journal of Theoretical and Applied Finance* **14**(1) 137–162.
- [17] Choi, Sungyong, Andrzej Ruszczyński, Yao Zhao. 2011. A multiproduct risk-averse newsvendor with law-invariant coherent measures of risk. *Operations Research* **59**(2) 346–364.
- [18] Cormen, Thomas H., Clifford Stein, Ronald L. Rivest, Charles E. Leiserson. 2001. *Introduction to Algorithms*. McGraw-Hill Higher Education.
- [19] Cotter, John, Kevin Dowd. 2006. Extreme spectral risk measures: An application to futures clearing-house margin requirements. *Journal of Banking & Finance* **30**(12) 3469 – 3485.
- [20] Cuoco, Domenico, Hua He, Sergei Isaenko. 2008. Optimal dynamic trading strategies with risk limits. *Operations Research* **56**(2) 358–368.
- [21] Detlefsen, Kai, Giacomo Scandolo. 2005. Conditional and dynamic convex risk measures. *Finance and Stochastics* **9**(4) 539–561.
- [22] Epstein, Larry G., Martin Schneider. 2003. Recursive multiple-priors. *Journal of Economic Theory* **113**(1) 1 – 31.
- [23] Föllmer, Hans, Irina Penner. 2006. Convex Risk Measures and the Dynamics of Their Penalty Functions. *Statistics & Decisions* **24**(1) 61–96.

- 
- [24] Föllmer, Hans, Alexander Schied. 2002. Convex measures of risk and trading constraints. *Finance and Stochastics* **6**(4) 429–447.
- [25] Föllmer, Hans, Alexander Schied. 2004. *Stochastic Finance in Discrete Time*. 2nd ed. Walter de Gruyter, Berlin, Germany.
- [26] Freund, Robert M., James B. Orlin. 1985. On the Complexity of Four Polyhedral Set Containment Problems. *Mathematical Programming* **33** 139–145.
- [27] Fujishige, Satoru. 2005. *Submodular Functions and Optimization, Annals of Discrete Mathematics*, vol. 58. 2nd ed. Elsevier.
- [28] Gilboa, Itzhak, David Schmeidler. 1989. Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* **18**(2) 141–153.
- [29] Goemans, Michel, Leslie Hall. 1996. The strongest facets of the acyclic subgraph polytope are unknown. William Cunningham, S. McCormick, Maurice Queyranne, eds., *Integer Programming and Combinatorial Optimization, Lecture Notes in Computer Science*, vol. 1084. Springer Berlin / Heidelberg, 415–429.
- [30] Hardy, Mary R., Julia L. Wirch. 2004. The iterated CTE: A dynamic risk measure. *North American Actuarial Journal* **8**(4) 62–75.
- [31] Horst, Reiner, Hoang Tuy. 2003. *Global Optimization*. 3rd ed. Springer.
- [32] Huang, Pu, Dan A. Iancu, Marek Petrik, Dharmashankar Subramanian. 2012. Static and dynamic Conditional Value at Risk. Working paper.
- [33] Huber, Peter J. 1981. *Robust Statistics*. John Wiley & Sons, Inc.
- [34] Iyengar, Garud N. 2005. Robust dynamic programming. *Mathematics of Operations Research* **30**(2) 257–280.
- [35] Jorion, Philippe. 2006. *Value at Risk*. 3rd ed. McGraw Hill.
- [36] Kusuoka, Shigeo. 2001. On Law Invariant Coherent Risk Measures. *Advances in Mathematical Economics* **3** 83–95.
- [37] Maccheroni, Fabio, Massimo Marinacci, Aldo Rustichini. 2006. Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica* **74**(6) 1447–1498.
- [38] Nilim, Arnab, Laurent El Ghaoui. 2005. Robust control of Markov decision processes with uncertain transition matrices. *Operations Research* **53**(5) 780–798.
- [39] Penner, Irina. 2007. Dynamic Convex Risk Measures: Time Consistency, Prudence and Sustainability. Ph.D. thesis, Humboldt-Universität zu Berlin.
- [40] Riedel, Frank. 2004. Dynamic coherent risk measures. *Stochastic Processes and their Applications* **112**(2) 185 – 200.
- [41] Rockafellar, R. Tyrrell, Stanislav Uryasev. 2002. Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance* **26**(7) 1443 – 1471.
- [42] Rockafellar, T. 1970. *Convex Analysis*. Princeton University Press.
- [43] Rockafellar, Tyrrell R., Stanislav Uryasev. 2000. Optimization of Conditional Value-at-Risk. *Journal of Risk* **2**(3) 21–41.
- [44] Roorda, Berend, J. M. Schumacher. 2008. How to apply tail value at risk over multiple time steps avoiding accumulation of conservatism and extra parameters. Extended abstract for the Bachelier Finance Society World Congress.
- [45] Roorda, Berend, J. M. Schumacher, Jacob Engwerda. 2005. Coherent acceptability measures in multi-period models. *Mathematical Finance* **15**(4) 589–612.
- [46] Roorda, Berend, J.M. Schumacher. 2007. Time consistency conditions for acceptability measures, with an application to tail value at risk. *Insurance: Mathematics and Economics* **40**(2) 209 – 230.
- [47] Ruszczyński, Andrzej. 2010. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming* **125** 235–261.
- [48] Ruszczyński, Andrzej, Alexander Shapiro. 2006. Conditional risk mappings. *Mathematics of Operations Research* **31**(3) 544–561.

- 
- [49] Ruszczyński, Andrzej, Alexander Shapiro. 2006. Optimization of convex risk functions. *Mathematics of Operations Research* **31**(3) 433–452.
- [50] Schmeidler, David. 1986. Integral representation without additivity. *Proceedings of the American Mathematical Society* **97** 255–261.
- [51] Schmeidler, David. 1989. Subjective probability and expected utility without additivity. *Econometrica* **57**(3) pp. 571–587.
- [52] Schrijver, Alexander. 2000. *Theory of Linear and Integer Programming*. 2nd ed. John Wiley & Sons.
- [53] Schrijver, Alexander. 2003. *Combinatorial Optimization: Polyhedra and Efficiency*. 1st ed. Springer.
- [54] Shapiro, Alexander. 2012. Minimax and risk averse multistage stochastic programming. *European Journal of Operational Research* **219**(3) 719 – 726.
- [55] Shapiro, Alexander, Darinka Dentcheva, Andrzej Ruszczyński. 2009. *Lectures on Stochastic Programming*. MPS / SIAM Series on Optimization, SIAM.
- [56] Tsanakas, Andreas. 2004. Dynamic capital allocation with distortion risk measures. *Insurance: Mathematics and Economics* **35**(2) 223 – 243.
- [57] Wang, Shaun S. 2000. A Class of Distortion Operators for Pricing Financial and Insurance Risks. *The Journal of Risk and Insurance* **67**(1) pp. 15–36.
- [58] Wirch, Julia Lynn, Mary R. Hardy. 1999. A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics* **25**(3) 337 – 347.