MATHEMATICS OF OPERATIONS RESEARCH
Vol. 40, No. 3, August 2015, pp. 655–682
ISSN 0364-765X (print) | ISSN 1526-5471 (online)

http://dx.doi.org/10.1287/moor.2014.0689

Tight Approximations of Dynamic Risk Measures

Dan A. Iancu
Graduate School of Business, Stanford University, Stanford, California 94305,
daniancu@stanford.edu

Marek Petrik, Dharmashankar Subramanian
IBM T.J. Watson Research Center, Yorktown Heights, New York 10598
{mpetrik@us.ibm.com, dharmash@us.ibm.com}

This paper compares two frameworks for measuring risk in a multiperiod setting. The first corresponds to applying a single coherent risk measure to the cumulative future costs, and the second involves applying a composition of one-step coherent risk mappings. We characterize several necessary and sufficient conditions under which one measurement always dominates the other and introduce a metric to quantify how close the two 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 upper bound can be exactly characterized and corresponds to a popular construction in the literature, whereas the tightest lower bound is not readily available. We show that testing domination and computing the approximation factors are generally NP-hard, even when the risk measures are comonotonic and law-invariant. However, we characterize conditions and discuss examples where polynomial-time algorithms are possible. One such case is the well-known conditional value-at-risk measure, which we explore in more detail. 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.

Keywords: coherent risk measures; dynamic consistency; tight approximations; submodular functions; polymatroids
MSC2000 subject classification: Primary: 68T37, 05B35, 52B40, 52B55, 52B11, 90C05, 90C27, 91B06, 91B30
OR/MS subject classification: Cost analysis/decision analysis: risk; mathematics: combinatorics, convexity, piecewise Linear; programming: integer
History: Received November 9, 2012; revised August 6, 2014. Published online in *Articles in Advance* January 6, 2015.

1. Introduction. In recent years, the framework of convex and coherent risk measures has emerged as an axiomatically justified and computationally tractable approach for quantifying the risk of an uncertain future outcome (Artzner et al. [4], Föllmer and Schied [23], Schmeidler [51]). It has also provided a strong bridge across several streams of research, including ambiguous representations of preferences in economics (Gilboa and Schmeidler [26], Schmeidler [52], Epstein and Schneider [20], Maccheroni et al. [36]); axiomatic treatments of market risk in financial mathematics (Artzner et al. [4], Föllmer and Schied [22]); actuarial science (Wirch and Hardy [59], Wang [58], Acerbi [2], Kusuoka [34], Tsanakas [57]); operations research (Ben-Tal and Teboulle [9], Ruszczyński and Alexander [49]); and statistics (Huber [31]). Our goal in the present paper is to take this axiomatic framework as given and analyze two distinct ways of using it to quantify risk in dynamic decision settings.

To fix ideas, let us consider a dynamic setting with a finite set of stages $t \in \{0, 1, \ldots, T\}$ and a scenario tree representation of uncertainty, where Ω_t is the set of nodes at stage t, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_T$ denotes the natural filtration generated by the elementary events in $\{\Omega_t\}_{t\in\{0,\ldots,T\}}$, and $\mathcal{X}_t$ is the space of functions that are $\mathcal{F}_t$-measurable.

In this setting, consider evaluating the risk of a single terminal cost $X \in \mathcal{X}_T$ from the perspective of an intermediate stage $t \in \{0, \ldots, T\}$. A first approach might be to apply a "desirable" functional to the cost, conditional on the available information. More formally, the risk in X evaluated at time t would be given by $\mu^i(X \mid i)$, where $i \in \Omega_t$ denotes the state of the world at time t and $\mu^i\colon \mathcal{X}_T \to \mathbb{R}$ is a *risk measure* constrained to obey particular desirable axioms. In practice, often the same risk measure $\mu^i = \mu$ is used, resulting in a compact and unified representation of risk preferences. The approach also has one other key advantage: with convex risk measures, *static* decisions that affect the cost X can be efficiently computed by combining simulation procedures with convex optimization (Rockafellar and Uryasev [43], Ruszczyński and Shapiro [50]). This has led to a wide adoption of the approach in practice (Jorion [33]), as well as in several academic papers (see, e.g., Basak and Shapiro [8], Cuoco et al. [17], and references therein).

The above "static" paradigm, however, is known to suffer from the serious shortcoming of generating potentially inconsistent preferences over risk profiles in time, whereby a decision maker might deem a terminal cost X riskier than Y in every state of the world at some time $t+1$ but nonetheless consider Y riskier than X at time t. This has been criticized from an axiomatic perspective because it is a staple of irrational behavior (Epstein and Schneider [20], Roorda et al. [46], Boda and Filar [11], Artzner et al. [5]). 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 [20], Ruszczyński and Alexander [49], Nilim and El Ghaoui [37], Iyengar [32]). To prevent such undesirable effects, additional conditions must be imposed on the risk measurement process at distinct time periods. More formally, if $\mu_t\colon \mathscr{X}_T \to \mathscr{X}_t$ denotes the risk mapping used at time t, with components given by $(\mu^i)_{i\in\Omega_t}$, then the axiom[1] of *time* (or *dynamic*) *consistency* asks that $\{\mu_t\}_{t=0}^{T-1}$ should satisfy the condition

$$\forall\, t \in \{0, \ldots, T-1\},\ \ \forall\, X, Y \in \mathscr{X}_T,\ \ \mu_{t+1}(X) \ge \mu_{t+1}(Y) \quad \text{implies} \quad \mu_t(X) \ge \mu_t(Y). \tag{1}$$

This requirement has 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, any $\{\mu_t\}_{t=0}^{T-1}$ satisfying property (1) can be written as

$$\mu_t(X) = \rho_{t+1}\big(\rho_{t+2}(\cdots(\rho_T(X))\cdots)\big), \tag{2}$$

where $\rho_{t+1}\colon \mathscr{X}_{t+1} \to \mathscr{X}_t$, $\forall\, t \in \{0, \ldots, T-1\}$ are single-period conditional risk mappings (see, e.g., Epstein and Schneider [20], Riedel [40], Artzner et al. [5], Detlefsen and Scandolo [18], Roorda et al. [46], Cheridito et al. [14], Roorda and Schumacher [45], Penner [38], Föllmer and Penner [21], Ruszczyński [48], and references therein). Apart from yielding consistent preferences, this compositional form also allows a recursive estimation of the risk and an application of the Bellman optimality principle (Nilim and El Ghaoui [37], Iyengar [32], Ruszczyński and Alexander [49]). This has led to its adoption in several application areas, such as actuarial science (Hardy and Wirch [28], Brazauskas et al. [13]); inventory management (Ahmed et al. [3]); commodity operations (Devalkar et al.[19]); hydro-thermal scheduling (Philpott et al. [39]); financial engineering (Lin and Peña [35], Rudloff et al. [47]); control and dynamical systems (Chow and Pavone [15]); and others.

From a pragmatic perspective, however, the compositional form entails a significantly more complicated risk assessment than the former (potentially inconsistent) approach. A decision maker needs to specify single-period conditional risk mappings for every future stage; furthermore, even if these corresponded to the same risk measure ρ, the result of the composition would no longer be easily interpretable and would bear no immediate relation to the original ρ. As an example, when ρ corresponds to the well-known Conditional (or Average)-Value-at-Risk (AVaR), the overall μ_t would correspond to the so-called "iterated conditional tail expectation (CTE)" (Hardy and Wirch [28], Brazauskas et al. [13], Roorda and Schumacher [45]), which does not lend itself to the same simple interpretation as AVaR. This has been acknowledged as one of the main problems in adopting the approach in practice (Rudloff et al. [47]). 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, for the iterated CTE, one is taking tail conditional expectations of quantities that are already tail conditional expectations. This has been recognized informally in the literature by Roorda and Schumacher [44, 45], who proposed new notions of time consistency avoiding the issue but without establishing formally if or to what degree the conservatism is 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 former in a way that tightly approximates the latter. This issue is relevant when considering dynamic decision problems under risk (Rudloff et al. [47]), but it seems to have been largely ignored in the literature. Most papers start with the premise of a given time-consistent risk measure as the objective in an application and characterize the resulting optimal policies through dynamic programming arguments (see, e.g., Ahmed et al. [3], Shapiro [55], Devalkar et al.[19], Lin and Peña [35], Chow and Pavone [15]).

With this motivation, the goal of this paper is to mathematically characterize the relation and trade-offs between the two measurement approaches above and to provide guidelines for constructing and estimating safe counterparts of one from the other. Our contributions are as follows.

- We provide several equivalent necessary and sufficient conditions that guarantee when a time consistent risk measure μ_C always over- (or under-) estimates risk as compared with an inconsistent measure μ_I. We argue that iterating the same μ_I does not necessarily over- (or under-) estimate risk as compared to a single static application of μ_I, and this is true even in the case considered in Roorda and Schumacher [44, 45]. 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.

[1] We note that there are several notions of time consistency in the literature (see Penner [38], Acciaio and Penner [1], Roorda and Schumacher [45] 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.

- We formalize the problem of characterizing and computing the smallest α_{μ_C,μ_I} and α_{μ_I,μ_C} such that $\mu_C \le \mu_I \le \alpha_{\mu_C,\mu_I} \cdot \mu_C$ and $\mu_I \le \mu_C \le \alpha_{\mu_I,\mu_C} \cdot \mu_I$, respectively. The smallest such factors, $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$, provide a compact notion of how closely a given μ_I can be multiplicatively approximated through lower (respectively, upper) bounding consistent measures μ_C, respectively. Since in practice μ_I may be far easier to elicit from observed preferences or to estimate from empirical data, characterizing and computing $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$ can be seen as the first step toward *constructing* the time-consistent risk measure μ_C that is "closest" to a given μ_I.
- Using results from the theory of submodularity and matroids, we specialize our results to the case when μ_I and μ_C are both comonotonic risk measures. We show that computing $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$ 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 μ_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 μ_I and show that it corresponds to a construction suggested in several papers in the literature (Epstein and Schneider [20], Roorda et al. [46], Artzner et al. [5], Shapiro [55]) that involves "rectangularizing" the set of probability measures corresponding to μ_I. This yields not only the smallest possible $\alpha^\star_{\mu_I,\mu_C}$ but also the uniformly tightest upper bound among all coherent upper bounds.
- We summarize results from our companion paper (Huang et al. [30]), which applies the ideas derived here to the specific case when both μ_I and μ_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 multiperiod problems. We give an exact analytical characterization for the tightest uniform upper bound to μ_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 multiperiod models.

The rest of the paper is organized as follows. Section 2 briefly reviews the necessary background in static risk measures and formalizes the precise 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 corresponding multiplicative scaling factors. Section 4 discusses 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 future directions.

1.1. Notation. With $i < j$, we use $[i, j]$ to denote the index set $\{i, \ldots, j\}$. For a vector $\mathbf{x} \in \mathbb{R}^n$ and $i \in \{1, \ldots, n\}$, we use x_i to denote the i-th component of $\mathbf{x}$. For a set $S \subseteq \{1, \ldots, n\}$, 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. The terms $\pi(S)$ or $\sigma(S)$ designate one particular such permutation, with $\pi(i)$ denoting the element of S appearing in the i-th position under permutation π.

We use Δ^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 $\operatorname{ext}(P)$ to denote the set of its extreme points.

Throughout the exposition, we adopt the convention that $0/0 = 0$.

2. Background in risk measures. Precise problem statement. The present section briefly introduces the class of risk measures examined in the paper, reviews the two approaches for measuring dynamic risk outlined in the introduction, and formally states and discusses the main questions addressed in this work.

2.1. Probabilistic model. Our notation and framework are closely in line with that of Shapiro et al. [56], to which we direct the reader for more details. For simplicity, we work with a scenario tree representation of the uncertainty space, where $t \in [0, T]$ denotes the time, and Ω_t is the set of nodes at stage $t \in [0, T]$. We denote by $\mathcal{C}_i$ the set of children of node $i \in \Omega_t$ and by $\mathcal{D}_i$ the set of all leaves descending from node i.[2] We also define $\mathcal{D}_U \stackrel{\text{def}}{=} \bigcup_{i\in U} \mathcal{D}_i$ for any set $U \subseteq \Omega_t$.

[2] More formally, with $\mathcal{D}_i = \{i\}$, $\forall\, i \in \Omega_T$, we recursively define $\mathcal{D}_i \stackrel{\text{def}}{=} \bigcup_{j\in\mathcal{C}_i} \mathcal{D}_j$, $\forall\, i \in \bigcup_{t=0}^{T-1} \Omega_t$.

We let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_T$ denote the natural filtration generated by the elementary events in $\{\Omega_t\}_{t\in[0,T]}$ and construct a probability space $(\Omega_T, \mathcal{F}_T, \mathbb{P})$ by introducing a *reference measure* $\mathbb{P} \in \Delta^{|\Omega_T|}$ satisfying[3] $\mathbb{P} > 0$. We let $\mathcal{X}_t$ denote the space of all functions $X_t\colon \Omega_t \to \mathbb{R}$ 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 $\mathbf{X}_t \in \mathbb{R}^{|\Omega_T|}$ the vector of induced scenario-values, and we identify the expectation of X_t with respect to $\mathbf{q} \in \Delta^{|\Omega_t|}$ as the scalar product $\mathbf{q}^T\mathbf{X}_t$. To simplify notation, we also identify any function $f\colon \mathcal{X}_{t+1} \to \mathcal{X}_t$ with a set of $|\Omega_t|$ functions and write $f \equiv (f_i)_{i\in\Omega_t}$, where $f_i\colon \mathbb{R}^{|\Omega_{t+1}|} \to \mathbb{R}$. 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 $f_i\colon \mathbb{R}^{|\mathcal{C}_i|} \to \mathbb{R}$.

2.2. Static risk measures. As stated in the introduction, our treatment aims to compare two frameworks for measuring dynamic risk, both of which are centered around the concept of *risk measures*. To formally introduce these concepts, consider a generic probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{X}$ be a linear space of random variables on Ω. The standard approach in the literature for assessing the risk in a *cost* $X \in \mathcal{X}$ is to use a functional $\mu\colon \mathcal{X} \to \mathbb{R}$ such that $\mu(X)$ represents the minimal reduction that makes the cost X acceptable. The following requirements, which we also adopt in the present paper, are typically imposed on μ.

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

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

[P3] *Convexity*. For any $X, Y \in \mathcal{X}_T$ 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 $X \in \mathcal{X}_T$ and any $\lambda \geq 0$, $\mu(\lambda X) = \lambda\mu(X)$.

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

For an in-depth discussion and critique of these axioms, we direct the reader to Artzner et al. [4], Föllmer and Schied [23], Schmeidler [51], and references therein. Any functional satisfying [P1–P2] is called a *risk measure*, and risk measures additionally satisfying [P3], [P3–P4], and [P3–P5] are called *convex*, *coherent* and *comonotonic*, respectively. The main focus of the present paper are coherent measures, but some results take a simpler form in the comonotonic case. To this end, we recall one of the main results in the literature, which is a universal representation theorem of coherent (comonotonic) risk measure (Schmeidler [51], Föllmer and Schied [23]).

Theorem 1. *A risk measure μ is coherent if and only if it can be represented as*

$$\mu(X) = \max_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_{\mathbb{Q}}[X] \tag{3}$$

for some $\mathcal{Q} \subseteq \Delta^{|\Omega|}$. Furthermore, if μ is comonotonic, then $\mathcal{Q} = \{\mathbb{Q} \in \Delta^{|\Omega|}\colon \mathbb{Q}(S) \leq c(S),\ \forall\, S \in \mathcal{F}\}$, 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 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 $c\colon 2^\Omega \to [0,1]$ 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: $c(\varnothing) = 0$ and $c(\Omega) = 1$
- submodular: $c(A\cap B) + c(A\cup B) \leq c(A) + c(B)$, $\forall\, A, B \subseteq \Omega$.

An example of such a risk measure, popular among academics and practitioners alike, is average value-at-risk at level $\varepsilon \in [0,1]$ ($\mathrm{AVaR}_\varepsilon$), also known as *conditional value-at-risk*, *tail value-at-risk* or *expected shortfall*. It is defined as

$$\mathrm{AVaR}_\varepsilon(X) \stackrel{\text{def}}{=} \frac{1}{\varepsilon}\int_{1-\varepsilon}^{1} \mathrm{VaR}_{1-t}(X)\,dt. \tag{4a}$$

where $\mathrm{VaR}_\varepsilon(X) \stackrel{\text{def}}{=} \inf\{m \in \mathbb{R}\colon \mathbb{P}[X - m > 0] \leq \varepsilon\}$ is the Value-at-Risk at level ε. Although AVaR is comonotonic, it is well known that VaR is not even convex since it fails requirement [P3].

[3] This is without loss of generality—otherwise, all arguments can be repeated on a tree where leaves with zero probability are removed.

2.3. Dynamic risk measures. Main problem statement. As discussed in the introduction, static (coherent or comonotonic) risk measures can be used as building blocks for measuring risk in dynamic settings. More formally, considering again the problem of assessing the risk of a terminal cost $Y \in \mathscr{X}_T$ from the perspective of time $t = 0$, two frameworks are typically considered:

- a potentially time-inconsistent measurement, which involves applying a single (coherent or comonotonic) risk measure $\mu_I: \mathscr{X}_T \to \mathbb{R}$ to assess the risk in Y, and
- a time-consistent measurement, which entails using a risk measure $\mu_C: \mathscr{X}_T \to \mathbb{R}$ obtained by composing one-step conditional risk mappings. More formally, $\mu_C(Y) = \mu_0(\mu_1(\cdots \mu_{T-1}(Y)\cdots))$, where $\mu_t: \mathscr{X}_{t+1} \to \mathscr{X}_t$, $\mu_{t+1} \equiv (\mu^i)_{i\in\Omega_t}$, and $\mu^i: \mathbb{R}^{|\mathscr{C}_i|} \to \mathbb{R}$ are (coherent or comonotonic) risk measures, for any $i \in \Omega_t$ and any $t \in [0, T-1]$.

The goal of the present paper is to take the first step toward better understanding the relation and trade-offs between these two ways of measuring risk. To this end, we address the following related problems.

Problem 1. Given μ_I and μ_C, test whether

$$\mu_C(Y) \le \mu_I(Y), \;\; \forall\, Y \in \mathscr{X}_T \quad \text{or} \quad \mu_I(Y) \le \mu_C(Y), \;\; \forall\, Y \in \mathscr{X}_T.$$

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

$$\text{if } \mu_C(Y) \le \mu_I(Y), \;\; \forall\, Y \qquad \text{then } \mu_I(Y) \le \alpha_{\mu_C,\mu_I} \cdot \mu_C(Y), \;\; \forall\, Y \in \mathscr{X}_T, \;\; Y \ge 0 \tag{5a}$$

$$\text{if } \mu_I(Y) \le \mu_C(Y), \;\; \forall\, Y \qquad \text{then } \mu_C(Y) \le \alpha_{\mu_I,\mu_C} \cdot \mu_I(Y), \;\; \forall\, Y \in \mathscr{X}_T, \;\; Y \ge 0. \tag{5b}$$

A satisfactory answer to Problem 1 would provide a test for whether one of the formulations is always over- or underestimating risk as compared to the other. A central issue of concern here is the computational tractability of these tests and how that relates to the representation of the risk measures μ_I and μ_C. To understand the relevance of Problem 2, note that the minimal factors $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$ satisfying (5a) and (5b), respectively, provide a compact notion of how closely μ_I can be approximated through lower or upper bounding consistent measures μ_C, respectively.[4] Since in practice it may be far easier to elicit or estimate a single static risk measure μ_I, characterizing and computing $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$ constitute the first step toward constructing the time-consistent risk measure μ_C that is "closest" to a given μ_I.

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

Remark 1. Requiring nonnegative Y in the text of Problem 2 might initially seem overly restrictive. However, if we insisted on $\mu_C(Y) \le \mu_I(Y)$ holding for *any* cost Y, and if μ_C, μ_I were allowed to take both positive and negative values, then the questions in Problem 2 would be meaningless, in that no feasible $\alpha_{\cdot,\cdot}$ would exist satisfying (5a) or (5b). To this end, we occasionally require the stochastic losses Y to be nonnegative. This is not too restrictive whenever a lower bound is available for Y;[5] furthermore, in specific applications (e.g., multiperiod inventory management; Ahmed et al. [3]), Y is the sum of intraperiod nonnegative costs, so the restriction is without loss.

3. Bounds for coherent and comonotonic risk measures. In this section, we seek answers to Problem 1 and Problem 2. To keep the discussion compact, we first treat an abstract setting of comparing two coherent measures on the same space of outcomes and then revisit and specialize the results to the comonotonic case in §3.2.

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

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

where $\mathscr{Q}_{1,2} \subseteq \Delta^{|\Omega|}$ are bounded polyhedra.[6] Our main focus is on (1) characterizing conditions such that $\mu_1(Y) \le \mu_2(Y)$, $\forall\, Y \in \mathscr{X}$, and (2) finding the smallest factor α such that $\mu_1(Y) \le \mu_2(Y) \le \alpha\mu_1(Y)$, $\forall\, Y \in \mathscr{X}$ $(Y \ge 0)$.

[4] A similar concept of inner and outer approximations by means of risk measures appears in Bertsimas and Brown [10]; however, the goal and analysis there are quite different since the question is to approximate one static risk measure by means of another.

[5] If Y_L is a lower bound, by using the cash-invariance property [P2], one could reformulate the original question with regard to the random loss $Y - Y_L$, which would be nonnegative.

[6] 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 all comonotonic risk measures, is simpler to describe, and is computationally advantageous because evaluating the risk measure entails solving a linear program (LP).

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

Proposition 1. *The inequality* $\mu_1(Y) \le \mu_2(Y)$, $\forall\, Y \in \mathcal{X}$ *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 $\mathcal{Q}_i$. For instance, 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 the other three possible cases (Freund and Orlin [24]).

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* $\mu_2(Y) \le \alpha\,\mu_1(Y)$, $\forall\, Y \in \mathcal{X}$.

Proof. For any $\alpha > 0$, the condition $\{\mu_2(Y) \le \alpha\,\mu_1(Y),\ \forall\, Y \in \mathcal{X}\}$ 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 §2.3, a sensible choice is to take $Y \ge 0$, 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 Appendix A for more details and references).

Definition 2. A nonempty 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 $0 \le \mathbf{y} \le \mathbf{x}$, we also have $\mathbf{y} \in Q$.

Definition 3. The *down-monotone closure* of a set $Q \subseteq \mathbb{R}^n_+$, denoted by $\mathrm{sub}(Q)$, is the smallest down-monotone set containing Q; i.e.,

$$\mathrm{sub}(Q) \stackrel{\text{def}}{=} \big\{\mathbf{x} \in \mathbb{R}^n_+ \colon \exists \mathbf{q} \in Q,\, \mathbf{x} \le \mathbf{q}\big\}.$$

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 (these scenarios are such that their probabilities may sum up to strictly less than 1). 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*

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

Proof. The inequality $\max_{\mathbf{q}\in\mathcal{Q}} \mathbf{q}^T\mathbf{Y} \le \max_{\mathbf{q}\in\mathrm{sub}(\mathcal{Q})} \mathbf{q}^T\mathbf{Y}$ follows because $\mathcal{Q} \subseteq \mathrm{sub}(\mathcal{Q})$. To prove the reverse, consider any $Y \ge 0$ and let $\mathbf{q}_1 \in \arg\max_{\mathbf{q}\in\mathrm{sub}(\mathcal{Q})} \mathbf{q}^T\mathbf{Y}$. By Definition 3, there exists $\mathbf{q}_2 \in \mathcal{Q}$ such that $\mathbf{q}_2 \ge \mathbf{q}_1 \ge 0$. Therefore, $\max_{\mathbf{q}\in\mathcal{Q}} \mathbf{q}^T\mathbf{Y} \ge \mathbf{q}_2^T\mathbf{Y} \ge \mathbf{q}_1^T\mathbf{Y} = \max_{\mathbf{q}\in\mathrm{sub}(\mathcal{Q})} \mathbf{q}^T\mathbf{Y}$. $\square$

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* $\mu_1(Y) \le \mu_2(Y)$, $\forall\, Y \in \mathcal{X}$ *holds if and only if* $\mathrm{sub}(\mathcal{Q}_1) \subseteq \mathrm{sub}(\mathcal{Q}_2)$.

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

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

($\Rightarrow$) Consider $\mathbf{q}_1 \in \mathrm{sub}(\mathcal{Q}_1)$. Then by Definition 3, there exists $\mathbf{q}_1' \in \mathcal{Q}_1$ such that $\mathbf{q}_1' \ge \mathbf{q}_1$. Since $\mathcal{Q}_1 \subseteq \mathcal{Q}_2$, we have $\mathbf{q}_1' \in \mathcal{Q}_2$, and therefore $\mathbf{q}_1 \in \mathrm{sub}(\mathcal{Q}_2)$.

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

$$\mathrm{sub}(\mathcal{Q}_1) \subseteq \mathrm{sub}(\mathcal{Q}_2) \ \Rightarrow\ \mathrm{sub}(\mathcal{Q}_1) \cap \Delta^{|\Omega|} \subseteq \mathrm{sub}(\mathcal{Q}_2) \cap \Delta^{|\Omega|} \ \Leftrightarrow\ \mathcal{Q}_1 \subseteq \mathcal{Q}_2. \quad \square$$

Figure 1 depicts examples of the sets $\mathcal{Q}_1, \mathcal{Q}_2$ and their down-monotone closures $\mathrm{sub}(\mathcal{Q}_1)$ and $\mathrm{sub}(\mathcal{Q}_2)$, respectively. Note that the conditions provided by Lemma 1 hold for *any* cost Y; i.e., nonnegativity 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 $\mathrm{sub}(\mathcal{Q}_{1,2})$, but not for $\mathcal{Q}_{1,2}$ (Freund and Orlin [24]).

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

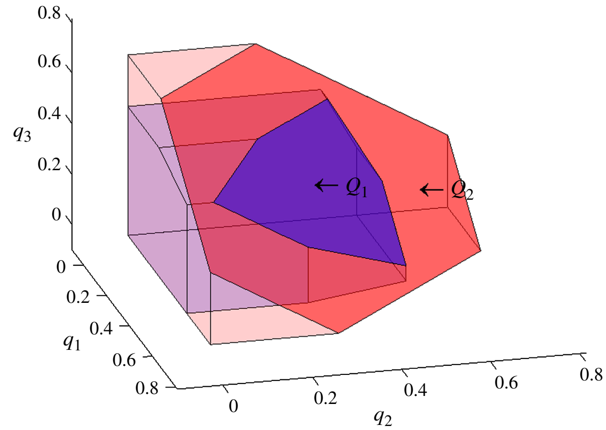

FIGURE 1. (Color online) Inclusion relation between $\mathcal{Q}_1, \mathcal{Q}_2$ (and the corresponding down-monotone closures, $\operatorname{sub}(\mathcal{Q}_1)$, and $\operatorname{sub}(\mathcal{Q}_2)$, respectively) that is equivalent to $\mu_1(Y) \le \mu_2(Y)$, $\forall\, Y \in \mathscr{X}_2$.

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

PROOF. By Proposition 2, $\mu_i(Y) = \max_{\mathbf{q} \in \operatorname{sub}(\mathcal{Q}_i)} \mathbf{q}^T Y$, for $i = 1, 2$. The claim then follows directly from Goemans and Hall [27, Lemma 1, also see Theorem 10 in the appendix]. □

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

THEOREM 2. *Let $\alpha^\star_{\mu_1,\mu_2}$ denote the smallest value of α such that $\mu_2(Y) \le \alpha \cdot \mu_1(Y)$, $\forall\, Y \ge 0$.*
1. *If $\operatorname{sub}(\mathcal{Q}_1) = \{\mathbf{q} \in \mathbb{R}^n_+ \colon \mathbf{a}_i^T \mathbf{q} \le b_i,\ \forall\, i \in \mathcal{I}\}$, where $\mathbf{a}_i \ge \mathbf{0}$, $b_i \ge 0$, then*

$$\alpha^\star_{\mu_1,\mu_2} = \max_{i \in \mathcal{I}} \frac{\max_{\mathbf{q} \in \operatorname{sub}(\mathcal{Q}_2)} \mathbf{a}_i^T \mathbf{q}}{b_i} = \max_{i \in \mathcal{I}} \frac{\max_{\mathbf{q} \in \mathcal{Q}_2} \mathbf{a}_i^T \mathbf{q}}{b_i} = \max_{i \in \mathcal{I}} \frac{\mu_2(\mathbf{a}_i)}{b_i}. \tag{7}$$

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

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

PROOF. The first claim is a known result in combinatorial optimization—see Goemans and Hall [27, Theorem 2, Theorem 10 in the appendix] for a complete proof.

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

$$\begin{aligned} &\min\Big\{ t\colon \max_{\mathbf{Y} \ge 0} \Big[\max_{\mathbf{q} \in \mathcal{Q}_2} \mathbf{Y}^T \mathbf{q} - t \cdot \max_{\mathbf{q} \in \mathcal{Q}_1} \mathbf{Y}^T \mathbf{q} \Big] \le 0 \Big\} \\ &\quad = \min\Bigg\{ t\colon \max_{\mathbf{Y} \ge 0} \Bigg[\max_{\mathbf{q} \in \mathcal{Q}_2} \mathbf{Y}^T \mathbf{q} - t \cdot \min_{\substack{A^T \boldsymbol{\mu} = \mathbf{Y} \\ \boldsymbol{\mu} \ge 0}} \mathbf{b}^T \boldsymbol{\mu} \Bigg] \le 0 \Bigg\} = \min\Bigg\{ t\colon \max_{\substack{\mathbf{q} \in \mathcal{Q}_2 \\ \boldsymbol{\mu} \ge 0,\, A^T \boldsymbol{\mu} \ge 0}} (A\mathbf{q} - t\mathbf{b})^T \boldsymbol{\mu} \le 0 \Bigg\}. \end{aligned}$$

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 separable inner optimization problems with a joint single maximization, switching the order of the maximizations, and eliminating the variables $\mathbf{Y}$. □

The results in Theorem 2 give a direct connection between the problem of computing $\alpha^\star_{\mu_1,\mu_2}$ and the representations available for the sets $\mathcal{Q}_i$ and $\operatorname{sub}(\mathcal{Q}_i)$. More precisely,

- If a polynomially sized inequality description is available for $\operatorname{sub}(\mathcal{Q}_1)$, then $\alpha^\star_{\mu_1,\mu_2}$ can be obtained by solving the small number of LPs in (7). Every such LP essentially entails an evaluation of the risk measure μ_2, leading to an efficient overall procedure.

Table 1. Computational complexity for determining whether $\mu_2(Y) \le \alpha \cdot \mu_1(Y)$, $\forall\, Y \ge \mathbf{0}$ 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.

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

- If a compact inequality representation is available for $\mathscr{Q}_1$, then $\alpha^\star_{\mu_1,\mu_2}$ can be found by bisection search over $t \ge 0$, where in each step the bilinear program in (8) is solved. Since bilinear programs can be reformulated as integer programs (Horst and Tuy [29]), 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 $\mathrm{sub}(\mathscr{Q}_2) \subseteq \alpha \cdot \mathrm{sub}(\mathscr{Q}_1)$ are summarized in Table 1. When $\mathscr{Q}_2$ or $\mathrm{sub}(\mathscr{Q}_2)$ have polynomially sized vertex descriptions, the test simply requires checking containment for a finite set of points, and when $\mathrm{sub}(\mathscr{Q}_2)$ has a polynomially sized description, the results of Theorem 2 apply. We conjecture that all the remaining cases are NP-complete but do not pursue a formal complexity 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.

We conclude our general discussion by noting that the tightest scaling factor $\alpha^\star_{\mu_1,\mu_2}$ can also be used to directly reexamine 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 $\mu_2(Y) \le \mu_1(Y)$, $\forall\, Y \in \mathscr{X}$ holds if and only if $\alpha^\star_{\mu_1,\mu_2} \le 1$.*

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 μ_1 and μ_2, by determining the scaling factors $\alpha^\star_{\mu_1,\mu_2}$ and $\alpha^\star_{\mu_2,\mu_1}$, we can readily test domination and also approximate one measure by the other, as follows:

- If $\alpha^\star_{\mu_2,\mu_1} \le 1$, then $\mu_1(Y) \in [1/\alpha^\star_{\mu_1,\mu_2}, 1] \cdot \mu_2(Y)$, $\forall\, Y \ge 0$.
- If $\alpha^\star_{\mu_1,\mu_2} \le 1$, then $\mu_1(Y) \in [1, \alpha^\star_{\mu_2,\mu_1}] \cdot \mu_2(Y)$, $\forall\, Y \ge 0$.

3.1. Tightest time consistent and coherent upper bound. The results and exposition in the prior section only required that the coherent risk measures $\mu_{1,2}$ have polyhedral sets of representing measures $\mathscr{Q}_1, \mathscr{Q}_2$. In this section, we discuss some of these results in the context of §2.3—more precisely, we take μ_1 as the time-inconsistent risk measure μ_I, and μ_2 denotes the compositional measure μ_C.

Our goal is to show that for coherent μ_I, a complete characterization of the tightest possible uniform upper bound to μ_I is readily available and is given by a popular construction in the literature (Epstein and Schneider [20], Roorda et al. [46], Artzner et al. [5], Shapiro [55]). This not only yields the tightest possible factor $\alpha^\star_{\mu_I,\mu_C}$, but also considerably simplifies the test $\mu_I(Y) \le \mu_C(Y)$, $\forall\, Y$, for any coherent μ_C.

The next proposition introduces this construction for an arbitrary coherent measure μ.

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

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

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

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

As mentioned, this idea has already been considered in the literature, and several authors have recognized that it provides an upper bound to μ. It is known that $\hat{\mu}$ is time consistent and has a representation of the form $\hat{\mu}(Y) = \sup_{\mathbf{q}\in\hat{\mathscr{Q}}_\mu} \mathbf{q}^T\mathbf{Y}$, $\forall\, Y \in \mathscr{X}_T$, where the set $\hat{\mathscr{Q}}_\mu$ has a *product* or *rectangular* structure.[7] Furthermore, $\mathscr{Q} \subseteq \hat{\mathscr{Q}}_\mu$, and therefore $\mu(Y) \le \hat{\mu}(Y)$, $\forall\, Y \in \mathscr{X}_T$ (Epstein and Schneider [20], Roorda et al. [46], Shapiro [55]).

[7] Note that it is obtained by computing products of the sets $\hat{\mathscr{Q}}^i_\mu$ of single-step conditional probabilities, obtained by marginalization at each node in the tree. However, this rectangularization procedure generally does not preserve comonotonicity; i.e., $\hat{\mu}$ may not be comonotonic, even when μ is.

In view of this, we claim that $\hat{\mu}_I$ actually represents the *tightest* upper bound for μ_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 $\hat{\mu}_I$ 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 $\rho_t \equiv (\rho_t^i)_{i\in\Omega_{t-1}}$: $\mathcal{X}_t \to \mathcal{X}_{t-1}$ are given by*

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

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*

$$\mu_C(Y) \geq \hat{\mu}_I(Y), \quad \forall\, Y \in \mathcal{X}_T \quad \text{and} \quad \alpha^\star_{\mu_I,\hat{\mu}} \leq \alpha^\star_{\mu_I,\mu_C}. \tag{11}$$

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

$$\hat{\mathcal{Q}}^i_{\mu_I} \subseteq \mathcal{Q}^i_\rho, \quad \forall\, i \in \Omega_{t-1}, \quad \forall\, t \in [1, T]. \tag{12}$$

Proof. 1. Since μ_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 $\mathcal{Q}^i_\rho$ (see, e.g., Roorda et al. [46] or Föllmer and Schied [23]). Because of 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 (Epstein and Schneider [20], Roorda et al. [46], Artzner et al. [5]); i.e.,

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

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 $\hat{Q}^i_{\mu_I} \subseteq \mathcal{Q}^i_\rho$, $\forall\, i \in \Omega_{t-1}, \forall\, t$. This readily implies that $\hat{\mathcal{Q}}_{\mu_I} \subseteq \mathcal{Q}_C$, and hence $\hat{\mu}_I(Y) \leq \mu_C(Y)$, $\forall\, Y$. The inequality for the multiplicative factors $\alpha^\star_{\mu_I,\cdot}$ 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 $\hat{\mathcal{Q}}^i_{\mu_I} \subseteq \mathcal{Q}^i_\rho$ implies that $\hat{\mathcal{Q}}_{\mu_I} \subseteq \mathcal{Q}_C$, and since $\mu_I(Y) \leq \hat{\mu}_I(Y)$, we have $\mu_I(Y) \leq \mu_C(Y)$, $\forall\, Y$. □

This result is useful in several ways. First, it suggests that the tightest time-consistent, coherent upper bound for a given μ_I is $\hat{\mu}_I$. This not only yields the smallest possible multiplicative factor $\alpha^\star_{\mu_I,\cdot}$, but the upper-bound is *uniform*, i.e., for any loss Y. Also, $\alpha^\star_{\mu_I,\hat{\mu}_I}$ is a lower bound on the best possible $\alpha^\star_{\mu_I,\mu_C}$ when the consistent measures μ_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 $\hat{\mathcal{Q}}^i_{\mu_I}, \mathcal{Q}^i_\rho \subseteq \Delta^{|\mathcal{C}_i|}$. This will prove relevant in our subsequent analysis of the case of comonotonic risk measures.

3.2. The comonotonic case. The results introduced in §§3 and 3.1 simplify when the risk measures μ_I and μ_C are further restricted to be comonotonic. We discuss a model with $T = 2$, but the results readily extend to a finite number of stages (see §3.4).

We start by characterizing μ_I, with its set of representing measures $\mathcal{Q}_I$ and its down closure $\mathrm{sub}(\mathcal{Q}_I)$. 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,[8] a concept studied extensively in combinatorial optimization (see Appendix B, Fujishige [25], and Schrijver [54] for a comprehensive review).

Proposition 5. *Consider a naïve dynamic comonotonic risk measure μ_I: $\mathcal{X}_2 \to \mathbb{R}$, 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: $2^{|\Omega_2|} \to \mathbb{R}$ such that*

[8] We note that with the exception of the normalization requirement $c(\Omega) = 1$, which is unimportant for analyzing fundamental structural properties, the definition of a *Choquet capacity* is identical to that of a *rank function of a polymatroid* (Fujishige [25, Chapter 2]). Therefore, we use the two names interchangeably throughout the current paper.

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

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

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

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

Proof. By Theorem 1 for comonotonic risk measures, there exists a Choquet capacity c such that $\mathcal{Q}_I = \{\mathbf{q} \in \Delta^{|\Omega_2|}: \mathbf{q}(S) \le c(S), \ \forall\, S \subseteq \Omega_2\}$. 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 11 in Apppendix B). □

In particular, both sets $\mathcal{Q}_I$ and $\operatorname{sub}(\mathcal{Q}_I)$ are polytopes contained in the nonnegative orthant and generally described by exponentially many inequalities, one for each subset of the ground set Ω_2. However, evaluating the risk measure μ_I for a given $Y \in \mathcal{X}_2$ can be done in time polynomial in $|\Omega_2|$ by a simple greedy procedure (see Föllmer and Schied [23, Theorem 12 in the appendix or Lemma 4.92]).

In view of the results in §3.1, one may also seek a characterization of the tightest upper bound to μ_I, i.e., $\hat{\mu}_I$, or of its set of representing measures $\hat{\mathcal{Q}}_{\mu_I}$. Unfortunately, this seems quite difficult for general Choquet capacities c, but is possible in at least one case of interest, namely, when μ_I is given by $\mathrm{AVaR}_\varepsilon$, which we discussed in our companion paper (Huang et al. [30]). However, 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 $\operatorname{sub}(\mathcal{Q}_C)$.

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

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

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

where $c_1: 2^{|\Omega_1|} \to \mathbb{R}$ *and* $c_{2|i}: \ |2^{|\mathcal{C}_i|} \to \mathbb{R}$, $\forall\, i \in \Omega_1$ *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*

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

$$= \{\mathbf{q} \in \mathbb{R}_+^{|\Omega_2|}: \exists \mathbf{p} \in \mathcal{P}_{c_1}: \mathbf{q}\,|_{\mathcal{C}_i} \in \mathcal{P}_{p_i \cdot c_{2|i}}, \quad \forall\, i \in \Omega_1\}$$

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 proof is technical and involves a repeated application of ideas similar to those in the proof of Proposition 5. Therefore, we relegate it to Appendix C. □

As expected, the set of product measures $\mathcal{Q}_C$ and its down-monotone closure $\operatorname{sub}(\mathcal{Q}_C)$ have a more complicated structure than $\mathcal{Q}_I$ and $\operatorname{sub}(\mathcal{Q}_I)$, respectively. However, they remain polyhedral sets, characterized by the base polytopes and polymatroids associated with the Choquet capacities c_1 and $c_{2|i}$. The inequality descriptions of $\mathcal{Q}_C$ and $\operatorname{sub}(\mathcal{Q}_C)$ 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 12 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 13 in the appendix or Schrijver [54] 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 μ_I and μ_C, as given by Proposition 5 and Proposition 6. Then*
1. *The extreme points of $\mathcal{Q}_I$ are given by*

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

where $\sigma \in \Pi(\Omega_2)$ is any permutation of the elements of Ω_2.
2. *The extreme points of $\mathcal{Q}_C$ are given by*

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

where $\pi \in \Pi(\Omega_1)$ is any permutation of the elements of Ω_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 13.

Part (2) follows by a repeated application of Theorem 13 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

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

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

3.3. Computing the optimal bounds $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$. With the representations provided above, we now derive one of our main technical results, establishing a method for computing the tightest multiplicative bounds for comonotonic risk measures.

Theorem 3. *For any pair of risk measures μ_I and μ_C as introduced in §3.2,*

$$\alpha^\star_{\mu_C,\mu_I} = \max_{\mathbf{q} \in \mathrm{sub}(\mathcal{Q}_I)} \max_{S \subseteq \Omega_1} \frac{\sum_{i \in S} \max_{U \subseteq \mathcal{C}_i} (\mathbf{q}(U))/(c_{2|i}(U))}{c_1(S)}, \tag{16}$$

$$\alpha^\star_{\mu_I,\mu_C} = \max_{\mathbf{q} \in \mathrm{sub}(\mathcal{Q}_C)} \max_{S \subseteq \Omega_2} \frac{\mathbf{q}(S)}{c(S)}. \tag{17}$$

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

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

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

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

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

The second set of constraints implies that any feasible $\mathbf{p}$ satisfies

$$p_i \geq \frac{1}{\alpha} \max_{U \subseteq \mathcal{C}_i} \frac{\mathbf{q}(U)}{c_{2|i}(U)}, \quad \forall\, i \in \Omega_1.$$

Corroborated with the first set of constraints, this yields

$$\frac{1}{\alpha}\sum_{i\in S}\max_{U\subseteq\mathcal{C}_i}\frac{\mathbf{q}(U)}{c_{2|i}(U)} \le \sum_{i\in S} p_i \le c_1(S),\quad \forall\, S\subseteq\Omega_1 \quad\Leftrightarrow\quad \alpha \ge \max_{S\subseteq\Omega_1}\frac{\sum_{i\in S}\max_{U\subseteq\mathcal{C}_i}(\mathbf{q}(U))/(c_{2|i}(U))}{c_1(S)}.$$

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

The expression for $\alpha^{\star}_{\mu_I,\mu_C}$ is a direct application of the second part of Theorem 2, by identifying $\operatorname{sub}(\mathcal{Q}_1)$ with $\operatorname{sub}(\mathcal{Q}_I)$ and using the compact representation for $\operatorname{sub}(\mathcal{Q}_I)$ 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 $\operatorname{sub}(\mathcal{Q})$ 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. □

From Theorem 3, it can readily seen that when $\mu_C\le\mu_I$, the optimal $\alpha^{\star}_{\mu_C,\mu_I}$ 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 $\mu_I\le\mu_C$, the optimal $\alpha^{\star}_{\mu_I,\mu_C}$ 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 μ_I and μ_C (Proposition 5 and Proposition 6) have the properties*

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

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 Artzner et al. [4], which states that any risk measure μ should satisfy $\mu(Y)>0$ for any random cost $Y\ge 0$, $Y\neq 0$. In this case, the polytopes $\mathcal{Q}_I$ and $\mathcal{Q}_C$ are both full-dimensional (Balas and Fischetti [6]), which leads to finite minimal scalings.

As suggested in our general exposition at the beginning of §3, determining the optimal scaling factors $\alpha^{\star}_{\mu_C,\mu_I}$ and $\alpha^{\star}_{\mu_I,\mu_C}$ also leads to direct conditions for determining whether μ_C lower bounds μ_I or vice versa. The following corollary states these in terms of optimization problems.

Corollary 3. *For any pair of risk measures μ_I and μ_C as introduced in §3.2,*

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

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

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

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

Proof. The proof follows from Corollary 2, by recognizing that the condition $\mu_C\le\mu_I$ is equivalent to setting $\alpha^{\star}_{\mu_I,\mu_C}\le 1$ (and similarly for the reverse inequality and $\alpha^{\star}_{\mu_C,\mu_I}$). The formulas in Theorem 3 then immediately yield the desired conclusions. □

The results in Corollary 3 are stated in terms of nontrivial 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 μ_I and μ_C as introduced in §3.2,*

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

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

where $(s_1,\ldots,s_{|\Omega_1|})$ denotes any permutation of the elements of Ω_1, and $U_i\subseteq\mathcal{C}_i$ for any $i\in\Omega_1$.

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

$$c\left(\bigcup_{i\in S} \mathcal{C}_i\right) \le c_1(S), \quad \forall\, S \subseteq \Omega_1,$$

$$\frac{c(U)}{c(U) + 1 - c(\Omega_2 \backslash \mathcal{C}_i \cup U)} \le c_{2,i}(U), \quad \forall\, U \subseteq \mathcal{C}_i, \quad \forall\, i \in \Omega_1.$$

Proof. The proof is slightly technical, so we defer it to Appendix C. □

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 $\mu_C \le \mu_I$, and $\mathcal{O}(|A| \cdot 2^{\max_{i\in A} |\mathcal{C}_i|})$ for $\mu_I \le \mu_C$, respectively, where $A \stackrel{\text{def}}{=} \bigcup_{t\in[0,T-1]} \Omega_t$. 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 §4.1 and the example in §4.3.

We note that the reason the conditions for $\mu_I \le \mu_C$ 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 $\mu_I \le \mu_C$ can be done by separately examining conditions at each node of the scenario tree.

3.4. Multistage 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 μ_I with a time-consistent, comonotonic $\mu_C \stackrel{\text{def}}{=} \mu_1 \circ \mu_2 \circ \cdots \circ \mu_T$. The former is exactly characterized by Proposition 5, whereas the representation for the latter can be summarized in the following extension of Proposition 6.

Proposition 8. *Consider a time-consistent, comonotonic risk measure μ_C. Then*
1. *There exists $\mathcal{Q}_C \subseteq \Delta^{|\Omega_T|}$ such that $\mu_C(Y) \stackrel{\text{def}}{=} \max_{\mathbf{q}\in\mathcal{Q}_C} \mathbf{q}^T\mathbf{Y}$, $\forall\, Y \in \mathscr{X}_2$.*
2. *The set of measures $\mathcal{Q}_C$ is given by*

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

where $c_{t|i}\colon 2^{|\mathcal{C}_i|} \to \mathbb{R}$ 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* $\operatorname{sub}(\mathcal{Q}_C)$ *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 3 for the optimal multiplicative factors to a multiperiod setting, as follows.

Theorem 4. *For any comonotonic measure μ_I and time-consistent comonotonic measure μ_C,*

$$\alpha^\star_{\mu_C,\mu_I} = \max_{\mathbf{q}\in\operatorname{sub}(\mathcal{Q}_I)} \max_{S\subseteq\Omega_1} \frac{\sum_{i\in S} z_1(i,\mathbf{q})}{c_1(S)}, \tag{19}$$

where $\mathbf{z}_T(i,\mathbf{q}) \stackrel{\text{def}}{=} q_i$, $\forall\, i \in \Omega_T$, and

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

Also,

$$\alpha^\star_{\mu_I,\mu_C} = \max_{\mathbf{q}\in\operatorname{sub}(\mathcal{Q}_C)} \max_{S\subseteq\Omega_T} \frac{\mathbf{q}(S)}{c(S)}. \tag{20}$$

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

The proof follows analogously to that of Theorem 3 by using the expressions for $\mathrm{sub}(\mathcal{Q}_I)$ and $\mathrm{sub}(\mathcal{Q}_C)$ provided by Proposition 5 and Proposition 8, respectively, to analyze the conditions $\mathrm{sub}(\mathcal{Q}_I) \subseteq \alpha \cdot \mathrm{sub}(\mathcal{Q}_C)$ or vice versa. We omit it for brevity.

By comparing (23) and (20) with their two-period analogues in (16) and (17), respectively, it is interesting to note that the complexity of the formulation for $\alpha^\star_{\mu_I,\mu_C}$ remains the same, whereas the optimization problems yielding $\alpha^\star_{\mu_C,\mu_I}$ get considerably more intricate. Section 4 contains a detailed analysis of the computational complexity surrounding these problems.

For completeness, we remark that direct multiperiod counterparts for Corollary 3 and Corollary 4 can be obtained by recognizing that $\mu_C(Y) \le \mu_I(Y)$, $\forall\, Y$ is equivalent to $\alpha^\star_{\mu_I,\mu_C} \le 1$ and by using the results in Theorem 4 and Lemma 2 to simplify the latter conditions. We do not include these extensions because of 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 $\alpha^\star_{\mu_I,\mu_C}$ and $\alpha^\star_{\mu_C,\mu_I}$ 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 μ_I, and how closely do they approximate the original measure? This section addresses such questions in detail.

4.1. Computational complexity. As argued in §3, computing the optimal scaling factors $\alpha^\star_{\mu_I,\mu_C}$ and $\alpha^\star_{\mu_C,\mu_I}$ 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 restricting attention to a special subclass of comonotonic measures, known as *distortion measures*. These measures correspond to Choquet capacities that are uniquely determined by a concave distortion of the underlying probability of events; i.e.,

$$c(S) = \Psi(\mathbb{P}(S)), \quad \forall\, S \in \mathcal{F}, \tag{21}$$

where $\Psi\colon [0,1] \to [0,1]$ is a concave, nondecreasing function satisfying $\Psi(0) = 0$ and $\Psi(1) = 1$. Such measures are known to satisfy additional axiomatic properties (such as law invariance) and have been studied extensively in the literature (see, e.g., Kusuoka [34], Acerbi [2]).

To prove hardness, we use a reduction from the NP-hard SUBSET-SUM problem (Cormen et al. [16]) which is defined as follows.

Definition 4 (SUBSET-SUM). Given a set of integers $\{k_1, k_2, \ldots, k_m\}$, is there a subset that sums to s?

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

Theorem 5. *Consider two arbitrary distortion risk measures $\mu_{1,2}\colon \mathcal{X}_1 \to \mathbb{R}$. Then it is NP-hard to decide if $\alpha^\star_{\mu_2,\mu_1} \ge \gamma$, for any $\gamma \ge 0$. The problem remains NP-hard even when $\mu_2(Y) \le \mu_1(Y)$, for all $Y \in \mathcal{X}_1 (Y \ge 0)$.*

Proof. 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

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

and $c_i(S) = \Psi_i(\mathbb{P}(S))$, where $\Psi_i\colon [0,1] \to [0,1]$ are concave, increasing functions satisfying $\Psi_i(0) = 0$, $\Psi_i(1) = 1$. Because both $\mathrm{sub}(\mathcal{Q}_1)$ and $\mathrm{sub}(\mathcal{Q}_2)$ are polymatroids and downward monotone, the second result in Theorem 3 can be further simplified to

$$\alpha^\star_{\mu_2,\mu_1} = \max_{S\subseteq\Omega_1} \frac{c_1(S)}{c_2(S)}. \tag{22}$$

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

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

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 (22) is upper bounded as

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

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) nondecreasing on the interval $[0, s/K)$ and nonincreasing 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 (22) is $\sqrt{K/s}$. Finally, the result also holds when $\mu_2(Y) \le \mu_1(Y)$ since our choice already has $c_2(S) \le c_1(S)$ for all $S \subseteq \Omega_1$, which implies $\mu_2(Y) \le \mu_1(Y)$. □

The complexity of computing the optimal scalings $\alpha^\star_{\mu_I,\mu_C}$ and $\alpha^\star_{\mu_C,\mu_I}$ readily follows as a direct corollary of Theorem 5.

Corollary 5. *Under a fixed $T \ge 1$ and for any given $\gamma \ge 0$, it is NP-complete to decide whether $\alpha^\star_{\mu_C,\mu_I} \ge \gamma$ for an arbitrary inconsistent comonotonic measure μ_I and a consistent comonotonic measure μ_C. The result remains true even when μ_C and μ_I are such that $\mu_C(Y) \le \mu_I(Y)$ for all $Y \in \mathscr{X}_T$ $(Y \ge 0)$. Similarly, it is NP-complete to decide whether $\alpha^\star_{\mu_I,\mu_C} \ge \gamma$, and the result remains true even when $\mu_I(Y) \le \mu_C(Y)$, $\forall\, Y \in \mathscr{X}_T$ $(Y \ge 0)$.*

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 5 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. □

Corollary 5 argues that computing the optimal scaling factors for arbitrary distortion risk measures cannot be done in polynomial time. Although the NP-hardness may be somewhat disappointing, solving the two optimization problems in Theorem 3 is nonetheless clearly preferable to simply examining all possible values of Y.

Although the problem of computing the scaling factors is hard for general distortion measures, polynomial-time algorithms are possible when the representations of $\mathscr{Q}_I, \mathscr{Q}_C$ or $\mathrm{sub}(\mathscr{Q}_I), \mathrm{sub}(\mathscr{Q}_C)$ fall in the tractable cases discussed in Table 1 of §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 $\mathscr{Q}_I$ is available, and problem (23) can be solved in time polynomial in $|\Omega_T|$, under oracle access to the Choquet capacities $c_{t|i}$ yielding the measure μ_C.

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

Proof. Consider the specialization of (23) for a fixed $\mathbf{q} \in \mathscr{Q}_I$:

$$\alpha^\star_{\mu_C,\mu_I} = \max_{\mathbf{q}\in \mathrm{sub}(\mathscr{Q}_I)} \max_{S\subseteq\Omega_1} \frac{\mathbf{z}_1(U,\mathbf{q})}{c_1(S)}, \tag{23}$$

where $\mathbf{z}_T(i,\mathbf{q}) \stackrel{\text{def}}{=} q_i$, $\forall\, i \in \Omega_T$, and

$$z_t(i,\mathbf{q}) \stackrel{\text{def}}{=} \max_{U\subseteq\mathcal{C}_i} \frac{\mathbf{z}_{t+1}(U,\mathbf{q})}{c_{t+1|i}(U)}, \quad \forall\, t \in [1, T-1], \quad \forall\, i \in \Omega_t.$$

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

$$z_t(i,\mathbf{q}) = \max_{U\subseteq\mathcal{C}_i} \frac{\mathbf{z}_{t+1}(U,\mathbf{q})}{c_{t+1|i}(U)} = \min\big\{l \in \mathbb{R}\colon\, l\cdot c_{t+1|i}(U) - \mathbf{z}_{t+1}(U,\mathbf{q}) \ge 0,\ \forall\, U \subseteq \mathcal{C}_i\big\}.$$

For any l, the constraint $l\cdot c_{t+1|i}(U) - \mathbf{z}_{t+1}(U,\mathbf{q}) \ge 0$, $\forall\, U \subseteq \mathcal{C}_i$ 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 (Schrijver [54]). □

The result above is slightly stronger than what Table 1 suggests since the representation of $\mathscr{Q}_C \equiv \mathscr{Q}_1$ 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, we now consider several constructions suggested in the literature. The starting point is typically a single risk measure $\mu_I\colon \mathscr{X}_2 \to \mathbb{R}$, 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 μ_I. The questions we would like to address here are which of these measures lower bound or upper bound the inconsistent evaluation μ_I and what can be said about the relative tightness of the various formulations.

To construct dynamically consistent measures by composing μ_I, we must first specify the conditional one-step risk mappings corresponding to μ_I, formally denoted by ${\mu_I}^1: \mathscr{X}_1 \to \mathbb{R}$ and ${\mu_I}^2: \mathscr{X}_2 \to \mathscr{X}_1$. This is most natural when μ_I is a *distortion* risk measure. Recall from our brief discussion in §4.1 that any such measure is uniquely specified by the concave function Ψ 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 Ψ to suitable conditional probabilities. More precisely, ${\mu_I}^1$ and ${\mu_I}^{2|i}$ are the distortion risk measures corresponding to the Choquet capacities:

$$c_1\colon 2^{\Omega_1} \to \mathbb{R},\, c_1(S) = \Psi\Bigg(\sum_{i\in S}\mathbb{P}(\mathcal{C}_i)\Bigg),\quad \forall\, S \subseteq \Omega_1,$$

$$c_{2\,|\,i}\colon 2^{\mathcal{C}_i} \to \mathbb{R},\, c_{2|i}(U_i) = \Psi\bigg(\frac{\mathbb{P}(U_i)}{\mathbb{P}(\mathcal{C}_i)}\bigg),\quad \forall\, U_i \subseteq \mathcal{C}_i,\quad \forall\, i \in \Omega_1.$$

The conditional risk mappings ${\mu_I}^1$ and ${\mu_I}^2$ can then be used to define dynamic time-consistent risk measures. Some examples often considered in the literature include

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

where $\mathbb{E}$ denotes the expectation operator, and max is the worst-case operator. When 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 ${\mu_I}^2$. (All the proofs in this section make use of the results in §3; we choose to relegate them to the appendix since they rather are technical and not very insightful.)

4.2.1. Time-consistent lower bounds derived from a given μ_I. We begin by discussing two choices for lower-bounding consistent risk measures derived from μ_I. The following proposition formally establishes the first relevant result.

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

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

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 (Föllmer and Schied [23]). The same remains true in dynamic settings, provided that μ_I is applied in a single time step and expectation operators are applied in all other stages.

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

1. For a given μ_I, is it true that $(\mu_I\circ\mathbb{E})(Y) \le (\mathbb{E}\circ\mu_I)(Y)$, $\forall\, Y\in\mathscr{X}_2$ (or vice versa)?
2. Is it true that $\alpha^\star_{\mu_I\circ\mathbb{E},\,\mu_I} \ge \alpha^\star_{\mathbb{E}\circ\mu_I,\,\mu_I}$ *for any* distortion measure μ_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 1. 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 $\mathbb{R}^{|\Omega_2|}$):

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

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

The second question can always be answered for a *specific* μ_I by calculating the optimal scalings, so it really makes sense when posed for *all* risk measures. Unfortunately, our computational experiments show that any one of the scaling factors can be better than the other, so neither of the two claims holds. 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 suggests that the two lower bounds result in *equal* tightness of approximation in certain cases of interest.

Theorem 6. *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 μ_I, we have*

$$\alpha^\star_{\mu_I\circ\mathbb{E},\,\mu_I} = \alpha^\star_{\mathbb{E}\circ\mu_I,\,\mu_I} = N\cdot\max\bigg\{\frac{\Psi(1/N^2)}{\Psi(1/N)}, \frac{\Psi(2/N^2)}{\Psi(2/N)}, \ldots, \Psi(1/N)\bigg\}.$$

4.2.2. Time-consistent upper bounds derived from a given μ_I. In an analogous fashion to the previous discussion, one can ask what time-consistent *upper bounds* can be derived from a distortion measure μ_I. A natural supposition, analogous to the results of §4.2.1, may be that $\mu_I \circ \max$ and $\max \circ \mu_I$ are upper bounds to μ_I since max is the most conservative risk mapping possible (Föllmer and Schied [23]). 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 measure μ_I, and the risk measures $\mu_I \circ \max$ and $\max \circ \mu_I$, where* max *denotes the conditional worst-case operator. Then*
(i) *For any cost $Y \in \mathscr{X}_2$, $\mu_I(Y) \le (\mu_I \circ \max)(Y)$.*
(ii) *There exists a choice of μ_I and of random costs $Y_{1,2} \in \mathscr{X}_2$ such that $(\max \circ \mu_I)(Y_1) < \mu_I(Y_1)$ and $(\max \circ \mu_I)(Y_2) > \mu_I(Y_2)$.*

The result in Proposition 10 also suggests that upper bounds to μ_I can be derived by composing μ_I with more conservative mappings in later time periods. This intuition is sharpened in §4.3 and our companion paper (Huang et al. [30]), which show that when $\mu_I = \mathrm{AVaR}_\varepsilon$, all upper bounds of the form $\mathrm{AVaR}_\varepsilon \circ \mathrm{AVaR}_\gamma$ must have $\gamma \le \varepsilon$, 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 μ_I, one can also turn to the question of comparing the resulting scaling factor $\alpha^\star_{\mu_I, \mu_I \circ \max}$ with the factors of the previous section, namely, $\alpha^\star_{\mu_I \circ \mathbb{E}, \mu_I}$ or $\alpha^\star_{\mathbb{E} \circ \mu_I, \mu_I}$. 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.

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

$$\alpha^\star_{\mu_I, \mu_I \circ \max} = \max\left\{ \frac{\Psi(1/N)}{\Psi(1/N^2)}, \frac{\Psi(2/N)}{\Psi(2/N^2)}, \ldots, \frac{1}{\Psi(1/N)} \right\}.$$

Corroborating this result with the expression in Theorem 6 for $\alpha^\star_{\mu_I \circ \mathbb{E}, \mu_I}$, one can readily find simple examples of distortions Ψ such that either the latter or the former scaling factor is smaller.

An opinion often held among practitioners, and informally argued in the literature (Roorda and Schumacher [44, 45]) 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 μ_I. For instance, if $\mu_I = \mathrm{AVaR}$—the case considered in Roorda and Schumacher [45]—the compositional measure corresponds to the so-called "iterated 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 2 (ITERATED AVaR). Consider a uniform scenario tree (i.e., $|\Omega_1| = |\mathscr{C}_i| = 4$, $\forall\, i \in \Omega_1$) and a uniform reference measure. Furthermore, consider the risk measure $\mu_I \equiv \mathrm{AVaR}_{3/4}$, and the following two costs (specified as real vectors in $\mathbb{R}^{|\Omega_2|}$, with components split in the four subtrees of stage $T = 2$):

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

When $M > -1$, it can be readily checked[9] that $\mu_I(X) = 1 > (\mu_I \circ \mu_I)(X) = (8 - M)/9$, and $(\mu_I \circ \mu_I)(Y) = 1 > \mu_I(Y) = (3 - M)/4$.

We direct the interested reader to our companion paper (Huang et al. [30]), 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.

4.3. The case of AVaR. Several of the results in 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 $\alpha^\star_{\mu_C, \mu_I}$ and $\alpha^\star_{\mu_I, \mu_C}$ and for testing $\mu_I(Y) \le \mu_C(Y)$ or vice versa. Furthermore, the question of designing the measures μ_C providing the tightest lower or upper approximations to a given μ_I can also be more satisfactorily addressed.

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

The case is discussed at length in our companion paper (Huang et al. [30]), 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 ($|\mathcal{C}_i| = N$, $\forall\, i \in \bigcup_{t=0}^{T-1} \Omega_t$, and $\mathbb{P} = \mathbf{1}/N^T$), and the following choice of risk measures:

$$\mu_I = \mathrm{AVaR}_\varepsilon, \quad \varepsilon \in [1/N^T, 1] \tag{24a}$$

$$\mu_C = \mathrm{AVaR}_{\varepsilon_1} \circ \mathrm{AVaR}_{\varepsilon_2} \circ \cdots \circ \mathrm{AVaR}_{\varepsilon_T}, \quad \varepsilon_t \in [1/N, 1], \ \ \forall\, t \in [1, T]. \tag{24b}$$

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

In this setup, we can revisit our main results in Theorem 3 and provide the following expressions for the tightest factors $\alpha^\star_{\mu_C,\mu_I}$ and $\alpha^\star_{\mu_I,\mu_C}$ for the case $T = 2$.

Theorem 7. *Consider a case $T = 2$, and the pair of risk measures in (24a) and (24b). Then*

$$\alpha^\star_{\mu_C,\mu_I} = \begin{cases} \max\left\{N\varepsilon_1, \dfrac{\varepsilon_1\varepsilon_2}{\varepsilon}, N\varepsilon_2\right\}, & \varepsilon \le \dfrac{1}{N} \\[2ex] \max\left\{\dfrac{\varepsilon_1}{\varepsilon}, f(N, \varepsilon, \varepsilon_2)\right\}, & \varepsilon > \dfrac{1}{N}, \end{cases} \tag{25a}$$

$$\alpha^\star_{\mu_I,\mu_C} = \max\left\{1, \frac{\varepsilon}{\varepsilon_1\ \varepsilon_2}\right\}, \tag{25b}$$

where $f(N, \varepsilon, \varepsilon_2)$ is an explicit analytical function. Furthermore, the result for $\alpha^\star_{\mu_I,\mu_C}$ 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 $\mu_C(Y) \le \mu_I(Y)$, $\forall\, Y$ (or vice versa) since the latter conditions are equivalent to $\alpha^\star_{\mu_I,\mu_C} \le 1$ (respectively, $\alpha^\star_{\mu_C,\mu_I} \le 1$). This leads to the following simple tests.

Corollary 6. *Consider the pair of risk measures in (24a) and (24b). Then*

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

$$\varepsilon_1\varepsilon_2 \ge \varepsilon; \tag{26}$$

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

$$\varepsilon_1 \le \max\left(\frac{1}{N}, \varepsilon\right) \quad \text{and} \quad \varepsilon_2 \le \max\left(\frac{1}{N}, N\varepsilon - N + 1\right). \tag{27}$$

Furthermore, (26) remains true under an arbitrary scenario tree and reference measure $\mathbb{P}$.

The latter result confirms the observation in Example 2 that the iterated AVaR, i.e., $\mu_C = \mathrm{AVaR}_\varepsilon \circ \mathrm{AVaR}_\varepsilon$, is generally neither an upper nor a lower bound to the inconsistent choice $\mu_I = \mathrm{AVaR}_\varepsilon$. By (27), $\varepsilon_1 = \varepsilon$ is always a feasible option, but one must take $\varepsilon_2 \le \max(1/N,\, N\varepsilon - N + 1)$. In fact, as argued in Huang et al. [30], most relevant choices of ε 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 μ_C that is the tightest approximation to a given $\mu_I = \mathrm{AVaR}_\varepsilon$. More precisely, one can characterize the choice of μ_C (i.e., levels $\varepsilon_{1,2}^{\mathrm{LB}}$) that results in the smallest possible factor $\alpha^\star_{\mu_C,\mu_I}$ among all compositional AVaR that are lower bounds for $\mathrm{AVaR}_\varepsilon$ and similarly the values $\varepsilon_{1,2}^{\mathrm{UB}}$ yielding the smallest possible $\alpha^\star_{\mu_I,\mu_C}$ among all upper-bounding compositional AVaRs. The optimal choices satisfy several interesting properties:

- For values of ε that are common in financial applications, i.e., satisfying $\varepsilon \le 1/N$ (Jorion [33]), the optimal $\alpha^\star_{\mu_C,\mu_I}$ is obtained by taking $\varepsilon_1^{\mathrm{LB}} = \varepsilon_2^{\mathrm{LB}} = \sqrt{\varepsilon}$, corresponding to an iterated AVaR measure.
- The optimal $\alpha^\star_{\mu_I,\mu_C}$ requires choosing $\varepsilon_1^{\mathrm{UB}} = \varepsilon$ and $\varepsilon_2^{\mathrm{UB}} = \max(1/N, N\varepsilon - N + 1)$. Typical values of ε used in practice would entail $\varepsilon_2^{\mathrm{UB}} = 1/N$, i.e., the worst-case scenario in the second stage.
- The optimally designed $\alpha^\star_{\mu_C,\mu_I}$ is *always* smaller than $\alpha^\star_{\mu_I,\mu_C}$, i.e., for every ε and N, which suggests that starting with an underestimating $\mathrm{AVaR}_\varepsilon$ results in tighter dynamically consistent approximations for AVaR.

The results discussed in Theorem 7 for $T = 2$ can also be (partially) extended to a case of an arbitrary T, which is summarized in the following claim.

THEOREM 8. *Consider an arbitrary T and the pair of risk measures in (24a) and (24b). Then*

1. *There is an algorithm that computes $\alpha^\star_{\mu_C,\mu_I}$ in time $\mathcal{O}(N^{T^2})$.*
2. *$\alpha^\star_{\mu_I,\mu_C} = \max\{1, \varepsilon/(\prod_{t=1}^T \varepsilon_t)\}$, and the expression remains valid for an arbitrary scenario tree and reference measure.*

It is interesting to note that computing $\alpha^\star_{\mu_I,\mu_C}$, and hence also testing $\mu_C \le \mu_I$, remains as easy for general T as for $T = 2$: an analytical expression is available that actually holds in considerably more general settings (arbitrary tree and reference measure). By contrast, computing $\alpha^\star_{\mu_C,\mu_I}$ and testing $\mu_I \le \mu_C$ 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 μ_I may lead to a tighter approximating μ_C, the gain does not come for free because the computation of the resulting $\alpha^\star_{\mu_C,\mu_I}$ is typically harder than that for $\alpha^\star_{\mu_I,\mu_C}$.

In a multiperiod setting, the question of designing the tightest possible lower-bounding approximation μ_C to a given μ_I becomes harder—even computing one scaling factor $\alpha^\star_{\mu_C,\mu_I}$ requires a polynomial-time algorithm. By contrast, a complete characterization of the tightest upper-bound μ_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 μ_I. This is summarized in the following result (for a proof, see Huang et al. [30]).

THEOREM 9. *Consider the risk measure $\mu_I = \mathrm{AVaR}_\varepsilon$, under an arbitrary reference measure $\mathbb{P}$, and the construction for the risk measure $\hat{\mu}_I$ characterized in Proposition 4. Then $\hat{\mu}_I = \mathrm{AVaR}_1 \circ \mathrm{AVaR}_2 \circ \cdots \circ \mathrm{AVaR}_T$, where $\mathrm{AVaR}_t = (\hat{\mu}_i)_{i\in\Omega_{t-1}}$, and*

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

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

This result, which holds under any reference measure $\mathbb{P}$, suggests that starting with $\mu_I = \mathrm{AVaR}_\varepsilon$ 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 γ_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 μ_I (see Lemma 2), this implies that the tightest possible choice for a compositional AVaR that upper bounds a given $\mathrm{AVaR}_\varepsilon$ is exactly $\widehat{\mathrm{AVaR}}_\varepsilon$. In a different sense, this also provides an instance when starting with a comonotonic (in fact, distortion) risk measure μ_I results in a comonotonic (distortion) risk measure $\hat{\mu}_I$, which furthermore belongs to the same class as μ_I.

Lastly, the theorem confirms that the best possible compositional AVaR that upper bounds $\mathrm{AVaR}_\varepsilon$ 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 $\gamma_i \ge \gamma_j$, $\forall\, j \in \mathcal{C}_i$, and once node i requires a worst-case operator, so will any descendant of i since $\mathbb{P}(\mathcal{D}_i) \ge \mathbb{P}(\mathcal{D}_j)$, $\forall\, j \in \mathcal{C}_i$. 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 $\mu_I = \mathrm{AVaR}_\varepsilon$.

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 $\mathrm{AVaR}_\varepsilon$ should use risk measurement processes that are compositions of increasingly conservative $\mathrm{AVaR}_\varepsilon$ 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.

Acknowledgments. The authors are grateful to Viswanath Nagarajan and Nikhil Bansal for the helpful discussions concerning the hardness proof in §4.1 and to the reviewing team for the insightful comments and suggestions, which have considerably improved the paper. Part of this research was conducted while Dan Iancu was with the Department of Mathematical Sciences of the IBM T.J. Watson Research Center, whose financial support is gratefully acknowledged.

Appendix A. Submissives, downward monotone closures, and antiblocking polyhedra. In the current section, we discuss the important notion of the *down monotone closure* of a polytope, also known as its *antiblocking polyhedron* or its *submissive*. Our exposition mostly follows Schrijver [53, Chapter 9], 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 *antiblocking type* if

$$Q \neq \varnothing, \quad Q \subseteq \mathbb{R}^n_+, \quad \text{and} \quad 0 \leq \mathbf{y} \leq \mathbf{x} \quad \text{and} \quad \mathbf{x} \in Q \text{ imply } \mathbf{y} \in Q.$$

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*

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

Proof. The proof follows closely from the definitions. We omit it here and direct the interested reader to Schrijver [53]. □

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

$$\operatorname{sub}(Q) \stackrel{\text{def}}{=} \left\{\mathbf{y} \in \mathbb{R}^n_+ \colon \exists \mathbf{x} \in Q, \mathbf{x} \geq \mathbf{y}\right\}. \tag{A1}$$

It can be easily checked that $\operatorname{sub}(Q) = (Q + \mathbb{R}^n_-) \cap \mathbb{R}^n_+$ and that $\operatorname{sub}(Q)$ is full-dimensional if and only if $Q \setminus \{\mathbf{x} \in \mathbb{R}^n \colon x_j = 0\} \neq \varnothing$, for all $j \in [1, n]$ (see Balas and Fischetti [6]). 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 Balas and Fischetti [6], Balas et al. [7], or Schrijver [53, Chapter 9] 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 [27] are references therein. The following result is relevant for our purposes.

Theorem 10. *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 $\mathbf{w} \in \mathbb{R}^n$,*

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

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

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

Proof. Part (1) is essentially Goemans and Hall [27, Lemma 1]. Since the latter reference omits a proof, we include one below, for completeness. "⇒" Follows trivially. "⇐" 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 $Q = \{\mathbf{x} \in \mathbb{R}^n_+ \colon \mathbf{a}_i^T\mathbf{x} \leq b_i, \ \forall i \in \mathcal{I}\}$, 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, $(1/\alpha)\max\{\mathbf{a}_j^T\mathbf{x} \colon \mathbf{x} \in P\} \geq (1/\alpha)\mathbf{a}_j^T\bar{\mathbf{x}} > b_j \geq \max\{\mathbf{a}_j^T\mathbf{x} \colon \mathbf{x} \in Q\}$.

Part (2) is exactly Goemans and Hall [27, Theorem 2], to which we direct the reader for a complete proof. □

The above result shows that $\alpha^\star$ can be $+\infty$, which is the case if Q has a strictly smaller dimension than P (in this case, some b_i are 0, whereas the corresponding d_i are strictly positive Schrijver [53]). However, if Q is full-dimensional, $\alpha^\star$ 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 Schrijver [54, Chapter 44, Vol. B] and Fujishige [25, Chapter 2, §3.3], to which we direct the interested reader for more information.

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

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

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 Ω, we define the following two polyhedra

$$\begin{aligned}
\mathcal{P}_c &\stackrel{\text{def}}{=} \left\{\mathbf{x} \in \mathbb{R}^{|\Omega|} \colon \mathbf{x} \geq 0, \mathbf{x}(S) \leq c(S), \ \forall S \subseteq \Omega\right\}, \\
\mathcal{EP}_c &\stackrel{\text{def}}{=} \left\{\mathbf{x} \in \mathbb{R}^{|\Omega|} \colon \mathbf{x}(S) \leq c(S), \ \forall S \subseteq \Omega\right\}.
\end{aligned} \tag{B1}$$

Note that $\mathcal{P}_c$ is nonempty if and only if $c \geq 0$ and that EP_c is nonempty if and only if $c(\varnothing) \geq 0$. 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, $c(\varnothing) = 0$, $c(\Omega) = 1$.

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, whereas 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 Schrijver [54]. *Let c be a submodular set function on Ω 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 most 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 $\mathcal{B}_c$,

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

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

Theorem 11. *For any submodular function c satisfying $c(\varnothing) = 0$,*
(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) *$\mathcal{P}_c = \operatorname{sub}(\mathcal{B}_c)$.*
(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) 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 $x_i = \mathbf{x}(\Omega) - \mathbf{x}(\Omega \backslash \{i\}) \geq c(\Omega) - c(\Omega \backslash \{i\})$.

(ii) "$\supseteq$" Follows trivially. "$\subseteq$" Consider any $\mathbf{y} \in \mathcal{EP}_c$. Without loss of generality,[10] assume $\mathbf{y}$ does *not* lie in the strict interior of $\mathcal{EP}_c$, and let $\mathcal{I}_{\mathbf{y}} \stackrel{\text{def}}{=} \{S \in \mathcal{F}\colon \mathbf{y}(S) = c(S)\}$ 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 11. 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 λ, $\mathbf{y}_\lambda \in \mathcal{EP}_c$. By making λ 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 $\operatorname{ext}(\mathcal{EP}_c) = \operatorname{ext}(\mathcal{B}_c)$.

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

(iv) Since λc 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 12 (Theorem 44.3, Corollaries 44.3(a, b) in Schrijver [54].). *Let $c\colon 2^\Omega \to \mathbb{R}$ be a submodular set function with $c(\varnothing) = 0$, and let $\mathbf{w} \in \mathbb{R}^{|\Omega|}_+$. The optimum solution of $\max_{\mathbf{x} \in \mathcal{EP}_c} \mathbf{w}^T\mathbf{x}$ is*

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

where $(s_1, \ldots, s_n)$ is a permutation of the elements of Ω such that $\mathbf{w}(s_1) \geq \mathbf{w}(s_2) \geq \cdots \mathbf{w}(s_n)$. 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 Schrijver [54]. $\square$

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

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

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

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

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

[10] 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 Fujishige [25, Theorem 3.22] and Schrijver [54, §44.6c]. □

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 $c(\varnothing) = 0$, $\mathcal{B}_c \subset \mathbb{R}^n_+$ 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 13) are nonnegative. "$\Rightarrow$" Consider any two sets $T \subset U \subseteq \Omega$, and take a chain of sets $S_1 \subset S_2 \subset \cdots \subset S_{|U\setminus T|}$ such that $S_1 = T$ and $S_{|U\setminus T|} = U$. By Theorem 13, 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 $\mu_t\colon \mathcal{X}_t \to \mathcal{X}_{t-1}$. 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*

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

where $c_1\colon 2^{|\Omega_1|} \to \mathbb{R}$ and $c_{2|i}\colon 2^{|\mathcal{C}_i|} \to \mathbb{R}$, $\forall\, i \in \Omega_1$ 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*

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

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 (Epstein and Schneider [20], Artzner et al. [5], Roorda et al. [46]), but we rederive it here together with the second claim to keep the paper self-contained. To this end, recall from §2.3 that μ_C can be written as $\mu_1 \circ \mu_2$, where $\mu_2 \equiv (\mu^i)_{i\in\Omega_1}$ and $\mu_1 : \mathcal{X}_1 \to \mathbb{R}$, as well as $\mu^i\colon \mathbb{R}^{|\mathcal{C}_i|} \to \mathbb{R}$, $\forall\, i \in \Omega_1$, are comonotonic risk measures. By Theorem 1, for any $X_1 \in \mathcal{X}_1$ and $X_2 \in \mathcal{X}_2$, we have

$$\mu_1(X_1) = \max_{\mathbf{p}\in\mathcal{Q}_1} \mathbf{p}^T\mathbf{X}_1, \quad \mathcal{Q}_1 \stackrel{\text{def}}{=} \left\{ \mathbf{p} \in \Delta^{|\Omega_1|}\colon \mathbf{p}(S) \leq c_1(S),\ \forall\, S \subseteq \Omega_1 \right\}, \tag{C1a}$$

$$\mu_2^i(X_2) = \max_{\mathbf{q}\in\mathcal{Q}_{2|i}} \mathbf{q}^T\mathbf{X}_2, \quad \mathcal{Q}_{2|i} \stackrel{\text{def}}{=} \left\{ \mathbf{q} \in \Delta^{|\Omega_2|}\colon \mathbf{q}(U) \leq c_{2|i}(U),\ \forall\, U \subseteq \mathcal{C}_i;\ \mathbf{q}|_{\Omega_2\setminus\mathcal{C}_i} = 0 \right\}. \tag{C1b}$$

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 $\mu_C(Y) = \max_{\mathbf{q}\in\tilde{Q}} \mathbf{q}^T\mathbf{Y}$, where $\tilde{Q}$ has the following product form structure (Shapiro et al. [56]):

$$\tilde{Q} = \left\{ \mathbf{q} \in \Delta^{|\Omega_2|}\colon \exists \mathbf{p} \in \mathcal{Q}_1, \exists \mathbf{q}^i \in \mathcal{Q}_{2|i},\ \forall\, i \in \Omega_1, \text{ such that } \mathbf{q} = \sum_{i\in\Omega_1} p_i \mathbf{q}^i \right\}. \tag{C2}$$

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

"$\subseteq$" Consider any $\mathbf{q} \in \tilde{Q}$, and let $\mathbf{p} \in \mathcal{Q}_1$ and $\mathbf{q}^i \in \mathcal{Q}_{2|i}$ denote the corresponding vectors in representation (C2). 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 $\mathbf{q}^i \stackrel{\text{def}}{=} \mathbf{q}_{\mathcal{C}_i}/p_i \in \mathcal{Q}_{2|i}$ (the only nonobvious 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, $\mathbf{q} = \sum_{i\in\Omega_1} p_i \mathbf{q}^i \in \tilde{Q}$. For completeness, we also note that $\mathbf{q}^i = \mathbf{q}_{\mathcal{C}_i}/p_i \in \mathcal{Q}_{2|i} \Leftrightarrow \mathbf{q}|_{\mathcal{C}_i}/p_i \in \mathcal{B}_{c_{2|i}} \Leftrightarrow \mathbf{q}|_{\mathcal{C}_i} \in \mathcal{B}_{p_i\cdot c_{2|i}}$, for any $i \in \Omega_1$ (by part (iv) of Theorem 11 in Appendix B).

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 Appendix B). As such, denote by $\mathcal{A}$ the set on the right of the equation.

"$\subseteq$" Consider an arbitrary $\mathbf{x} \in \operatorname{sub}(\mathcal{Q}_C)$. 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 $\mathbf{p} \in \mathcal{P}_{c_1} \equiv \operatorname{sub}(\mathcal{B}_{c_1})$, $\exists\, \bar{\mathbf{p}} \in \mathcal{B}_{c_1}$ such that $\bar{\mathbf{p}} \geq \mathbf{p} \geq 0$. Furthermore, $\mathbf{q}|_{\mathcal{C}_i} \in \mathcal{P}_{\bar{p}_i\cdot c_{2|i}} \equiv \operatorname{sub}(\mathcal{B}_{\bar{p}_i\cdot c_{2|i}})$, 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 $\mathbf{q} \in \operatorname{sub}(\mathcal{Q}_C)$. □

Corollary 4. *For any pair of risk measures μ_I and μ_C as introduced in §3.2,*
1. *The inequality $\mu_C(Y) \le \mu_I(Y)$, $\forall Y \in \mathscr{X}_2$ holds if and only if*

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

where $(s_1, \ldots, s_{|\Omega_1|})$ denotes any permutation of the elements of Ω_1, and $U_i \subseteq \mathcal{C}_i$ for any $i \in \Omega_1$.
2. *The inequality $\mu_I(Y) \le \mu_C(Y)$, $\forall Y \in \mathscr{X}_2$ holds if and only if*

$$c\left(\bigcup_{i\in S}\mathcal{C}_i\right) \le c_1(S), \quad \forall S \subseteq \Omega_1,$$

$$\frac{c(U)}{c(U)+1-c(\Omega_2\backslash\mathcal{C}_i\cup U)} \le c_{2,i}(U), \quad \forall U \subseteq \mathcal{C}_i, \quad \forall i \in \Omega_1.$$

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

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

$$\max_{\mathbf{q}\in \mathrm{ext}(\mathscr{Q}_C)} \mathbf{q}(S) \le c(S), \quad \forall S \subseteq \Omega_2.$$

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

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

where π is any permutation of Ω_1, and σ_ℓ 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 σ_ℓ yielding a maximal value of $\mathbf{q}(U_\ell)$ is always of the form $(\sigma(U_\ell), \sigma(\mathcal{C}_\ell\backslash U_\ell))$; i.e., it has the elements of U_ℓ in the first $|U_\ell|$ positions. This is because the functions $c_{2|\ell}$ are submodular so that $c(U_\ell) - c(\varnothing) \ge c(U_\ell \cup A) - c(A)$, for any $A \subseteq \mathcal{C}_\ell\backslash U_\ell$. With this recognition, the optimal permutations σ_ℓ always result in $\mathbf{q}(U_\ell) = [c_1(\bigcup_{k=1}^{\ell}\pi(k)) - c_1(\bigcup_{k=1}^{\ell-1}\pi(k))]c_{2|\ell}(U_\ell)$. 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 $\hat{\mu}_I$, the tightest possible coherent upper bound to μ_I. To this end, first recall the representation for $\mathscr{Q}_C$ in Proposition 6, pasted below for convenience:

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

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

$$\max_{\mathbf{q}\in\mathscr{Q}_I} \mathbf{q}\left(\bigcup_{i\in S}\mathcal{C}_i\right) \le c_1(S), \quad \forall S \subseteq \Omega_1, \tag{*}$$

$$\max_{\mathbf{q}\in\mathscr{Q}_I\colon \mathbf{q}(\mathcal{C}_i)\ne 0} \frac{\mathbf{q}(U)}{\mathbf{q}(\mathcal{C}_i)} \le c_{2|i}(U), \quad \forall U \subseteq \mathcal{C}_i, \quad \forall i \in \Omega_1. \tag{**}$$

We now argue that (*) and (**) are equivalent to the conditions in part (2). Recalling the description of $\mathscr{Q}_I$ in Proposition 5 and because any inequality $\mathbf{q}(S) \le c(S)$ is tight at some set S (also see Theorem 13), it can be seen that the maximum value of $\mathbf{q}(\bigcup_{i\in S}\mathcal{C}_i)$ in (*) is exactly $c(\bigcup_{i\in S}\mathcal{C}_i)$, 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. □

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*

$$\max_{\mathbf{q}\in\mathscr{Q}_I\colon\, \mathbf{q}(\mathcal{D}_i)\ne 0} \frac{\mathbf{q}(\mathcal{D}_U)}{\mathbf{q}(\mathcal{D}_i)} = \frac{c(\mathcal{D}_U)}{c(\mathcal{D}_U)+1-c(\Omega_T\backslash\mathcal{D}_i\cup\mathcal{D}_U)}. \tag{C3}$$

Proof. Since the problem on the left is a fractional linear program, the maximum is reached at an extreme point of $\mathscr{Q}_I$ (Boyd and Vandenberghe [12]). Note also that the objective is increasing in any q_j, $j \in \mathcal{D}_U$, and decreasing in any q_s, $s \in \mathcal{D}_i\backslash\mathcal{D}_U$.

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

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

let $\mathbf{v}^\sigma \in \operatorname{ext}(\mathcal{Q}_I)$ be the extreme point corresponding to $\sigma \in \Pi(\Omega_T)$. We claim that there exists an optimal solution in (C3) such that the permutation σ is of the form

$$\begin{aligned}\mathcal{D}_U &= \{\sigma(1), \cdots, \sigma(|\mathcal{D}_U|)\},\\ \mathcal{D}_i \backslash \mathcal{D}_U &= \{\sigma(|\Omega_T \backslash \mathcal{D}_i \cup \mathcal{D}_U| + 1), \ldots, \sigma(|\Omega_T|)\};\end{aligned} \tag{C4}$$

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

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

To see this, consider any permutation σ such that $\mathbf{v}^\sigma$ is optimal in (C3), yet there exist $j \in \mathcal{D}_U$ and $s \in \mathcal{D}_i \backslash \mathcal{D}_U$ 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 σ between $\bar{k}$ and k). Consider a new permutation π 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,

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

By a similar argument, $v_s^\pi \leq v_s^\sigma$. Furthermore, by construction, $v_r^\pi = v_r^\sigma$, $\forall\, r \in \mathcal{D}_i \backslash \{j, s\}$ 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 $\mathbf{v}^\pi(\mathcal{D}_i \backslash \mathcal{D}_U) \leq \mathbf{v}^\sigma(\mathcal{D}_i \backslash \mathcal{D}_U)$, 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 (without loss of generality) σ contains the elements of $\mathcal{D}_U$ before those of $\mathcal{D}_i \backslash \mathcal{D}_U$, a similar interchange argument can be done with respect to $\Omega_T \backslash \mathcal{D}_i$, to reach the conclusion (C4). 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 $\mu_I: \mathcal{X}_2 \to \mathbb{R}$, 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. 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 Ψ denote the distortion function corresponding to the (distortion) measure μ_I; i.e., the Choquet capacity is given by $c(S) = \Psi(\mathbb{P}(S))$, $\forall\, S \subseteq \Omega_2$, where $\Psi: [0, 1] \to [0, 1]$ is concave, nondecreasing, with $\Psi(0) = 0$, $\Psi(1) = 1$. Recall from §4.2 that the (compositional) risk measure $\mu_I \circ \mathbb{E}$ (or, more correctly, ${\mu_I}^1 \circ \mathbb{E}$) exactly corresponds to the following choice of Choquet capacities for the first and second stage, respectively:

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

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

Note that the same distortion function Ψ (yielding the risk measure μ_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

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

To see this, one can use the decreasing marginal returns property of Ψ; i.e.,

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

to argue that $(\Psi(\sum_{j=1}^{i} p_{\sigma(j)}) - \Psi(\sum_{j=1}^{i-1} p_{\sigma(j)}))/(p_{\sigma(i)}) \leq (\Psi(\sum_{j=1}^{i} u_{\sigma(j)}) - \Psi(\sum_{j=1}^{i-1} u_{\sigma(j)}))/(u_{\sigma(i)})$. 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

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

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

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

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

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

THEOREM 6. *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* μ_I, *we have*

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

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 μ_I is given by a concave distortion function $\Psi\colon [0,1]\to[0,1]$. To this end, by applying the result in Theorem 3, our goal is to argue that

$$\begin{aligned}\alpha^\star_{\mathbb{E}\circ\mu_I,\,\mu_I} &\stackrel{\text{def}}{=} \max_{\mathbf{q}\in\operatorname{ext}(\mathcal{Q}_I)}\max_{S\subseteq\Omega_1}\frac{\sum_{i\in S}\max_{U_i\subseteq\mathcal{C}_i}(\mathbf{q}(U_i))/(\Psi(\mathbb{P}(U_i)/\mathbb{P}(\mathcal{C}_i)))}{\sum_{i\in S}\mathbb{P}(\mathcal{C}_i)}\\ &= \max_{\mathbf{q}\in\operatorname{ext}(\mathcal{Q}_I)}\max_{S\subseteq\Omega_1}\frac{\sum_{i\in S}\max_{U_i\subseteq\mathcal{C}_i}(\mathbf{q}(U_i))/(\mathbb{P}(U_i)/\mathbb{P}(\mathcal{C}_i))}{\Psi(\sum_{i\in S}\mathbb{P}(\mathcal{C}_i))}\stackrel{\text{def}}{=}\alpha^\star_{\mu_I\circ\mathbb{E},\,\mu_I}.\end{aligned}\tag{C5}$$

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

LEMMA 4. *Consider the maximization problems yielding* $\alpha^\star_{\mathbb{E}\circ\mu_I,\,\mu_I}$ *in* (C5). *We claim the following*:
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*

$$\{\sigma(1),\ldots,\sigma(N)\} = \mathcal{C}_i,\tag{C6}$$

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* $i\in\Omega_1$,

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

4. $\alpha^\star_{\mathbb{E}\circ\mu_I,\,\mu_I} = \max_{i\in\Omega_1}\max_{U_i\subseteq\mathcal{C}_i}((\Psi(\mathbb{P}(U_i)))/(\mathbb{P}(\mathcal{C}_i)\Psi(\mathbb{P}(U_i)/\mathbb{P}(\mathcal{C}_i))))$.

PROOF OF LEMMA 4. Claim 1 follows from the mediant inequality. To see this, for a fixed $\mathbf{q}$, let

$$v_i \stackrel{\text{def}}{=} \max_{U_i\subseteq\mathcal{C}_i}\frac{\mathbf{q}(U_i)}{\Psi(\mathbb{P}(U_i)/\mathbb{P}(\mathcal{C}_i))},\quad \forall\, i\in\Omega_1,$$

and note that the maximum over $S\subseteq\Omega_1$ is achieved at any singleton $\{i\}\subseteq\arg\max\{v_\ell/\mathbb{P}(\mathcal{C}_\ell)\colon \ell\in\Omega_1\}$.

To see Claim 2, first recall that the set $\operatorname{ext}(\mathcal{Q}_I)$ corresponds to all possible permutations of $\mathcal{C}_2$ (Proposition 7). By Claim 1, since the inner maximum always occurs at a singleton $i^\star(\mathbf{q})$, the optimal $\mathbf{q}^\star$ must be such that components in $\mathcal{C}_{i^\star(\mathbf{q}^\star)}$ are "as large as possible." Because of the concavity of Ψ, this occurs when they appear in the first N positions in the permutation σ (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 $\mathbf{q}$. □

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

LEMMA 5. *Consider the maximization problems yielding* $\alpha^\star_{\mu_I\circ\mathbb{E},\,\mu_I}$ *in* (C5). *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* $\mathbf{q}^\star(S)\in\mathcal{Q}_I$ *corresponds to a permutation* $\sigma^S\in\Pi(\Omega_2)$ *such that*

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

In other words, the first $|S|$ *elements in the permutation* σ *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.*,

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

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

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

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 σ^S that have elements from distinct subtrees in the first $|S|$ components. For any such $\sigma(j)$, with $j \in \{1, \ldots, |S|\}$,

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

where $\sigma(j) \in \mathcal{C}_i$. By the concavity of Ψ, 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, $\alpha^\star_{\mathbb{E} \circ \mu_I, \mu_I} = \alpha^\star_{\mu_I \circ \mathbb{E}, \mu_I}$.

Proof of Theorem 6. By Lemma 4, $\alpha^\star_{\mathbb{E} \circ \mu_I, \mu_I} = \max_{i \in \Omega_1} \max_{U_i \subseteq \mathcal{C}_i} \Psi(\mathbb{P}(U_i))/(\mathbb{P}(\mathcal{C}_i)\Psi(\mathbb{P}(U_i))/(\mathbb{P}(\mathcal{C}_i)))$. For a uniform reference measure, because of the symmetry, this expression becomes

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

Similarly, by Lemma 5,

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

which becomes, under uniform reference measure,

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

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

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

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

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

Proof. (i) Let the Choquet capacity yielding the distortion measure μ_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

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

$$c_{2|i}\colon 2^{\mathcal{C}_i} \to \mathbb{R},\ c_{2|i}(U_i) = 1, \quad \forall U_i \neq \varnothing \subseteq \mathcal{C}_i, \quad \forall i \in \Omega_1.$$

The conditions to check from Corollary 4 are

$$\Psi\left(\mathbb{P}\left(\bigcup_{i \in S} \mathcal{C}_i\right)\right) \le \Psi\left(\sum_{i \in S} \mathbb{P}(\mathcal{C}_i)\right), \quad \forall S \subseteq \Omega_1,$$

$$\frac{\Psi(\mathbb{P}(U))}{\Psi(\mathbb{P}(U)) + 1 - \Psi(\mathbb{P}(\Omega_2 \backslash \mathcal{C}_i \cup U))} \le 1, \quad \forall U \subseteq \mathcal{C}_i, \quad \forall i \in \Omega_1.$$

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

(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 $\mu_I = \mathrm{AVaR}_{1/2}$, 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 $\mu_I(Y_1) = 0.44 > (\max \circ \mu_I)(Y_1) = 0.4$, but $\mu_I(Y_2) = 0.3 < (\max \circ \mu_I)(Y_2) = 1$. $\square$

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 μ_I, we have*

$$\alpha^\star_{\mu_I, \mu_I \circ \max} = \max\left\{ \frac{\Psi(1/N)}{\Psi(1/N^2)}, \frac{\Psi(2/N)}{\Psi(2/N^2)}, \ldots, \frac{1}{\Psi(1/N)} \right\}.$$

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

$$c_1\colon 2^{\Omega_1} \to \mathbb{R},\, c_1(S) = \Psi\left(\sum_{i\in S} \mathbb{P}(\mathcal{C}_i)\right) \equiv \Psi\left(\frac{|S|}{N}\right), \quad \forall\, S \subseteq \Omega_1,$$

$$c_{2|i}\colon 2^{\mathcal{C}_i} \to \mathbb{R},\, c_{2|i}(U_i) = 1, \quad \forall\, U_i \neq \varnothing \subseteq \mathcal{C}_i, \quad \forall\, i \in \Omega_1.$$

By Theorem 3, the optimal scaling factor is given by $\alpha^\star_{\mu_I,\mu_C} = \max_{\mathbf{q}\in\mathcal{Q}_C} \max_{S\subseteq\Omega_2} \mathbf{q}(S)/\Psi(|S|/N^2)$. Let us switch the order of the maximizations, and fix an arbitrary $S = \bigcup_{i\in\Omega_1} U_i \subseteq \Omega_2$. Using the representation of $\mathcal{Q}_C$ provided by Proposition 6, it can be readily seen that $\mathbf{q}(U_i) = 0$ if $U_i = \varnothing$, and $\mathbf{q}(U_i) \le p_i$, otherwise, where $\mathbf{p}(S) \le c_1(S)$, $\forall\, S \subseteq \Omega_1$. 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 $\alpha^\star_{\mu_I,\mu_C}$, immediately leads to the desired result.

References

[1] Acciaio B, Penner I (2011) Dynamic risk measures. Di Nunno G, Øksendal B, eds. *Advanced Mathematical Methods for Finance* (Springer, Berlin, Heidelberg), 1–34.
[2] Acerbi C (2002) Spectral measures of risk: A coherent representation of subjective risk aversion. *J. Banking Finance* 26(7):1505–1518.
[3] Ahmed S, Çakmak U, Shapiro A (2007) Coherent risk measures in inventory problems. *Eur. J. Oper. Res.* 182(1):226–238.
[4] Artzner P, Delbaen F, Eber J-M, Heath D (1999) Coherent measures of risk. *Math. Finance* 9(3):203–228.
[5] Artzner P, Delbaen F, Eber J-M, Heath D, Ku H (2007) Coherent multiperiod risk adjusted values and Bellman's principle. *Ann. Oper. Res.* 152:5–22.
[6] Balas E, Fischetti M (1996) On the monotonization of polyhedra. *Math. Programming* 78(7):59–84.
[7] Balas E, Bockmayr A, Pisaruk N, Wolsey L (2004) On unions and dominants of polytopes. *Math. Programming* 99(2):223–239.
[8] Basak S, Shapiro A (2001) Value-at-risk-based risk management: Optimal policies and asset prices. *Rev. Financial Stud.* 14(2):371–405.
[9] Ben-Tal A, Teboulle M (2007) An old-new concept of convex risk measures: The optimized certainty equivalent. *Math. Finance* 17(3):449–476.
[10] Bertsimas D, Brown DB (2009) Constructing uncertainty sets for robust linear optimization. *Oper. Res.* 57(6):1483–1495.
[11] Boda K, Filar J (2006) Time consistent dynamic risk measures. *Math. Methods Oper. Res.* 63(1):169–186.
[12] Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, Cambridge, UK).
[13] Brazauskas V, Jones BL, Puri ML, Zitikis R (2008) Estimating conditional tail expectation with actuarial applications in view. *J. Statist. Planning Inference* 138(11):3590–3604.
[14] Cheridito P, Delbaen F, Kupper M (2006) Dynamic monetary risk measures for bounded discrete-time processes. *Electronic J. Probab.* 11(3):57–106.
[15] Chow Y-L, Pavone M (2014) A unifying framework for time-consistent, risk-averse model predictive control: Theory and algorithms. *Proc. Amer. Control Conf.* (IEEE, Piscataway, NJ), 4204–4211.
[16] Cormen TH, Stein C, Rivest RL, Leiserson CE (2001) *Introduction to Algorithms* (McGraw-Hill Higher Education, New York).
[17] Cuoco D, He H, Isaenko S (2008) Optimal dynamic trading strategies with risk limits. *Oper. Res.* 56(2):358–368.
[18] Detlefsen K, Scandolo G (2005) Conditional and dynamic convex risk measures. *Finance Stochastics* 9(4):539–561.
[19] Devalkar S, Anupindi R, Sinha A (2012) Dynamic risk management of commodity operations: Model and analysis. Available at SSRN: http://ssrn.com/abstract=2432067.
[20] Epstein LG, Schneider M (2003) Recursive multiple-priors. *J. Econom. Theory* 113(1):1–31.
[21] Föllmer H, Penner I (2006) Convex risk measures and the dynamics of their penalty functions. *Statist. Decisions* 24(1):61–96.
[22] Föllmer H, Schied A (2002) Convex measures of risk and trading constraints. *Finance Stochastics* 6(4):429–447.
[23] Föllmer H, Schied A (2004) *Stochastic Finance in Discrete Time*. 2nd ed. (Walter de Gruyter, Berlin).
[24] Freund RM, Orlin JB (1985) On the complexity of four polyhedral set containment problems. *Math. Programming* 33(2):139–145.
[25] Fujishige S (2005) *Submodular Functions and Optimization,* 2nd ed. Annals of Discrete Mathematics, vol. 58 (Elsevier, Amsterdam).
[26] Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *J. Math. Econom.* 18(2):141–153.
[27] Goemans M, Hall L (1996) The strongest facets of the acyclic subgraph polytope are unknown. William C, McCormick S, Queyranne M, eds. *Integer Programming and Combinatorial Optimization,* Lecture Notes in Computer Science, Vol. 1084 (Springer, Berlin), 415–429.
[28] Hardy MR, Wirch JL (2004) The iterated CTE: A dynamic risk measure. *North Amer. Actuarial J.* 8(4):62–75.
[29] Horst R, Tuy H (2003) *Global Optimization*, 3rd ed. (Springer, Berlin).
[30] Huang P, Iancu DA, Petrik M, Subramanian D (2012) The price of dynamic inconsistency for distortion risk measures. arXiv:1106.6102v1.
[31] Huber PJ (1981) *Robust Statistics* (John Wiley & Sons, New York).
[32] Iyengar GN (2005) Robust dynamic programming. *Math. Oper. Res.* 30(2):257–280.
[33] Jorion P (2006) *Value at Risk*, 3rd ed. (McGraw Hill, New York).
[34] Kusuoka S (2001) On law invariant coherent risk measures. *Adv. Math. Econom.* 3(1):83–95.
[35] Lin Q, Peña JF (2012) Optimal trade execution with coherent dynamic risk measures. Available at SSRN: http://ssrn.com/abstract=2150878.
[36] Maccheroni F, Marinacci M, Rustichini A (2006) Ambiguity aversion, robustness, and the variational representation of preferences. *Econometrica* 74(6):1447–1498.
[37] Nilim A, El Ghaoui L (2005) Robust control of Markov decision processes with uncertain transition matrices. *Oper. Res.* 53(5):780–798.
[38] Penner I (2007) Dynamic convex risk measures: Time consistency, prudence and sustainability. Unpublished doctoral dissertation, Humboldt-Universität zu Berlin.
[39] Philpott A, de Matos V, Finardi E (2013) On solving multistage stochastic programs with coherent risk measures. *Oper. Res.* 61(4):957–970.
[40] Riedel F (2004) Dynamic coherent risk measures. *Stochastic Processes Their Appl.* 112(2):185–200.

[41] Rockafellar RT, Uryasev S (2002) Conditional value-at-risk for general loss distributions. *J. Banking Finance* 26(7):1443–1471.
[42] Rockafellar T (1970) *Convex Analysis* (Princeton University Press, Princeton, NJ).
[43] Rockafellar TR, Uryasev S (2000) Optimization of conditional value-at-risk. *J. Risk* 2(3):21–41.
[44] Roorda B, Schumacher JM (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 B, Schumacher JM (2007) Time consistency conditions for acceptability measures, with an application to tail value at risk. *Insurance: Math. Econom.* 40(2):209–230.
[46] Roorda B, Schumacher JM, Engwerda J (2005) Coherent acceptability measures in multiperiod models. *Math. Finance* 15(4):589–612.
[47] Rudloff B, Street A, Valladão DM (2012) Time consistency and risk averse dynamic decision models: Definition, interpretation and practical consequences. *Eur. J. Oper. Res.* 234(3):743–750.
[48] Ruszczyński A (2010) Risk-averse dynamic programming for Markov decision processes. *Math. Programming* 125(2):235–261.
[49] Ruszczyński A, Alexander S (2006a) Conditional risk mappings. *Math. Oper. Res.* 31(3):544–561.
[50] Ruszczyński A, Shapiro A (2006b) Optimization of convex risk functions. *Math. Oper. Res.* 31(3):433–452.
[51] Schmeidler D (1986) Integral representation without additivity. *Proceedings of the Americal Math. Soc.* 97(2):255–261.
[52] Schmeidler D (1989) Subjective probability and expected utility without additivity. *Econometrica* 57(3): 571–587.
[53] Schrijver A (2000) *Theory of Linear and Integer Programming*, 2nd ed. (John Wiley & Sons, Chichester, UK).
[54] Schrijver A (2003) *Combinatorial Optimization: Polyhedra and Efficiency* (Springer, Berlin).
[55] Shapiro A (2012) Minimax and risk averse multistage stochastic programming. *Eur. J. Oper. Res.* 219(3):719–726.
[56] Shapiro A, Dentcheva D, Ruszczyński A (2009) *Lectures on Stochastic Programming*, MPS/SIAM series on optimization (SIAM, Philadelphia).
[57] Tsanakas A (2004) Dynamic capital allocation with distortion risk measures. *Insurance: Math. Econom.* 35(2):223–243.
[58] Wang SS (2000) A class of distortion operators for pricing financial and insurance risks. *J. Risk Insurance* 67(1): 15–36.
[59] Wirch JL, Hardy MR (1999) A synthesis of risk measures for capital adequacy. *Insurance: Math. Econom.* 25(3):337–347.