manuscript

Robust CARA Optimization

Li Chen
Institute of Operations Research and Analytics, National University of Singapore
chen_l@u.nus.edu

Melvyn Sim
Department of Analytics & Operations, NUS Business School, National University of Singapore
dscsimm@nus.edu.sg

We propose robust optimization models and their tractable approximations that cater for ambiguity-averse decision makers whose underlying risk preferences are consistent with *constant absolute risk aversion* (CARA). Specifically, we focus on maximizing the worst-case expected exponential utility where the underlying uncertainty is generated from a set of stochastically independent factors with ambiguous marginals. To obtain computationally tractable formulations, we propose a hierarchy of approximations, starting from formulating the objective function as tractable concave functions in affinely perturbed cases, developing approximations in concave piecewise affinely perturbed cases, and proposing new multi-deflected linear decision rules for adaptive optimization models. We also extend the framework to address a multi-period consumption model. The resultant models would take the form of an exponential conic optimization problem (ECOP), which can be practicably solved using current off-the-shelf solvers. We present numerical examples including project management and multi-period inventory management with financing to illustrate how our approach can be applied to obtain high-quality solutions that could outperform current stochastic optimization approaches, especially in situations with high risk aversion levels.

Key words: robust optimization, constant absolute risk aversion, exponential cone programming

History: April 4, 2023.

1. Introduction

Optimization under uncertainty is a fundamental problem in operations research that can have significant practical impact. Unlike a deterministic optimization problem, the objective function in a decision model facing uncertainty is contingent on the preference of the decision maker. Indeed, decision preference concerning uncertainty can be associated with *risk* and *ambiguity*. In this paper, we refer to risk in situations where the true probability distribution, also known as physical probabilities (Anscombe and Aumann 1963), of the underlying random variables is known, while in ambiguity or *Knightian uncertainty* (Knight 1921), the true distribution would be unknown. For a random payoff with known distribution, risk aversion relates to the preference of a more predictable, but possibly lower payoff, over a highly unpredictable, but possibly higher payoff. However, if the true distribution is unknown, the actual payoff risk would also be unknown. Ambiguity aversion relates to the cautious behavior of evaluating the payoff risk using the worst distribution among an acceptable set of uncertain probabilities. Both risk and ambiguity aversion are sensible and relevant in decision making, not only for individuals or firms, but also for some government agencies where their policies could impact public health or safety.

In articulating the risk preference of the decision maker, we focus on the paradigm of expected utility introduced by Von Neumann and Morgenstern (1947). This framework has been widely accepted as the normative standard for rational agents in assessing uncertain payoffs, despite being contradicted by numerous experimental outcomes (*e.g.* Allais 1953). Various modifications are developed to explain the deficiencies, including rank dependent expected utility theory (Quiggin 1982), prospect theory (Kahneman and Tversky 1979), and bounded rationality (Simon 1955). In the expected utility framework, risk aversion is associated with an increasing and concave utility function. In particular, we will focus on the exponential utility function, which is uniquely associated with decision makers whose preferences are consistent with *constant absolute risk aversion* (CARA) (Arrow 1965, Pratt 1964). In such preferences, the risk tolerance level, which is defined by the magnitude of the ratio of the first to the second derivative of the utility function, is a constant and does not depend on the payoff amount. The risk tolerance level can also be interpreted as the payoff amount for which the decision maker would be roughly indifferent between accepting or rejecting a gamble involving a 50-50 chance of winning that amount and losing half that amount (see, *e.g.*, Delquié 2008). Because decision models with an exponential utility function are often analytically tractable, it is used as an adequate approximation of general utility functions (see Kirkwood 2004, for more details). For instance, the exponential utility preference preserves the structural properties of Markov decision process (Howard and Matheson 1972). In addition, if the distribution of the random payoffs is Gaussian, then the CARA preferences would be consistent with the mean-variance preferences (Markowitz 1952); this property has been used to reveal the high risk aversion of investors in the well-known *equity premium puzzle* of Mehra and Prescott (1985). As such, CARA preferences are commonly assumed in the literature of economics, finance and operations research (see, *e.g.*, Abbas and Howard 2015, Holmstrom and Milgrom 1991, Bouakiz and Sobel 1992, Veronesi 1999, Feng and Xiao 2008, Hall et al. 2015). In the literature review on the application of risk aversion, Corner and Corner (1995) note that the CARA utility function is about five times more commonly adopted than other types of utility functions combined.

Dantzig (1955) proposes the stochastic linear optimization model as a computational framework for optimization under uncertainty. A stochastic linear optimization model is typically a linear optimization problem; it has a risk-neutral objective function and the random variable is limited to discrete probability distribution with a modest number of scenarios. However, if the number of scenarios is infinite or exponential, solving the stochastic optimization model would be computationally intractable (Dyer and Stougie 2006, Hanasusanto et al. 2016). Nevertheless, we can still obtain approximate solutions via sample average approximation (SAA) approach, which is a popular randomized approximation technique that improves with the number of samples (see, *e.g.* Shapiro et al. 2014). For large sample size, many large-scale linear optimization techniques have also been developed for solving the stochastic optimization problem more efficiently (see, *e.g.,* Kall et al. 1994, Birge and Louveaux 2011, Prékopa 2013). Although we can extend this framework to an objective function based on the expected exponential utility, it would result in a nonlinear convex exponential conic optimization problem (ECOP), whose formulation size depends on the number of scenarios (see, *e.g.,* Dowson et al. 2020). Unlike the linear optimization models, such problems may not scale as well computationally with the number of samples; moreover, the quality of the approximation may depend on the risk tolerance level. Consequently, because of the computational limit on the number of samples, solutions obtained via SAA may suffer from the *optimizer's curse* and result in poor out-of-sample performance (Smith and Winkler 2006). Hence, apart from SAA approximation, there is a need to also consider other approximation approaches that would scale well computationally.

There is also a need to address ambiguity because in real world optimization problems under uncertainty, the true probability distribution of the underlying random variable is often unavailable. Gilboa and Schmeidler (1989) axiomatize the preference of an ambiguity-averse decision maker and propose the decision criterion that evaluates the worst-case expected utility over an ambiguity set of prior distributions, though the earliest application of such decision preference may be traced to Scarf (1957). Similar decision criteria have also been proposed in the *distributionally robust optimization (DRO)* within the operations research community (see, *e.g.*, Delage and Ye 2010, Esfahani and Kuhn 2018, Chen et al. 2020). While there are tractable distributionally robust optimization frameworks for piecewise linear utility functions (see, *e.g.*, Bertsimas et al. 2010, Wiesemann et al. 2014), they have yet been extended to exponential utility.

Summary of contributions

In this paper, we propose a tractable robust CARA optimization framework to address distributionally robust optimization problems with CARA preferences, applicable to various decision problems such as, *inter alia*, network lot-sizing, project management, and inventory management. We summarize our contribution as follows.

(i) We demonstrate that, unlike DRO problems with a risk-neutral objective, the robust CARA optimization problem with an affine payoff function is NP-hard even under a simple ambiguity set with mean and polytope support set. To alleviate the computational intractability while maintaining practical modeling, we focus on uncertainty generated from stochastically independent factors with ambiguous marginals and reformulate the problem as exponential conic optimization problems, which can be practically solved with off-the-shelf solvers.

(ii) We develop a tractable exponential conic optimization approximation for robust CARA optimization problems with concave piecewise affinely perturbed payoff functions, in contrast to existing literature that focuses on affine perturbations. We prove the associated properties and demonstrate its advantages over a Monte-Carlo approach.

(iii) We extend these approximations to address adaptive optimization by proposing a new piecewise affine decision rule called the multi-deflected linear decision rule (MLDR). We show that the MLDR outperforms existing decision rules in recourse approximation and achieves optimality under complete recourse condition with one recourse variable.

(iv) We also extend the MLDR to solve a multi-period consumption model while preserving non-anticipative requirements.

(v) We present numerical examples, including project management and multi-period inventory management with financing, to demonstrate the effectiveness of our tractable approximations in obtaining high-quality solutions that may outperform current stochastic optimization approaches, particularly in scenarios with high risk aversion levels.

Notations We use boldface lowercase letters to represent vectors such as $\boldsymbol{a}$, and calligraphic font to denote a set such as $\mathcal{Z}$. We denote by $\mathbb{R}$, $\mathbb{R}_+$ the set of all and non-negative real numbers, respectively. We denote by $[N] \triangleq \{1,2,...,N\}$ the set of positive running indices up to N. We denote by $|\mathcal{Z}|$ the cardinality of the set $\mathcal{Z}$. We use $\|\cdot\|$ to denote Euclidean norm of a vector. We denote by $\mathcal{P}_0(\mathcal{Z})$ the set of all probability distributions on support set $\mathcal{Z}$. We use tilde $(\tilde{\cdot})$ to denote uncertain parameters and use $\mathbb{P}$ to denote probability measure on sample space $\mathcal{Z}$. We denote by $\mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}]$ the expectation of $\tilde{\boldsymbol{z}}$ under probability distribution $\tilde{\boldsymbol{z}} \sim \mathbb{P}$. The inequality between two uncertain parameters $\tilde{t} \geq \tilde{v}$ describes state-wise dominance, *i.e.*, $\tilde{t}(\omega) \geq \tilde{v}(\omega)$ for all $\omega \in \Omega$. We denote $\mathcal{R}^{k,n}$ as the space of all measurable functions from $\mathbb{R}^k$ to $\mathbb{R}^n$ that are bounded on compact sets. We define plus function $(x)^+ \triangleq \max\{x, 0\}$.

2. Robust optimization with CARA preferences

We first focus on a decision model where we denote $\boldsymbol{x} \in \mathcal{X} \subseteq \mathbb{R}^{I_x}$ as a vector of *here-and-now* decision variables and $f(\boldsymbol{x}, \boldsymbol{z})$ being the payoff function in which the second argument, $\boldsymbol{z} \in \mathcal{Z} \subseteq \mathbb{R}^{I_z}$ represents the model's parameters that are subject to uncertainty. After the decision $\boldsymbol{x}$ has been made, the payoff function would be randomly perturbed by the uncertain parameters. We next present the following assumption on the model of uncertainty.

Assumption 1 **(Independent factors with ambiguous marginals)**. *We assume that the model's uncertainty is generated from a set of stochastically independent factors with ambiguous marginals, denoted by* $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \tilde{z}_2, \cdots, \tilde{z}_{I_z})$ *where* $\tilde{z}_j$, $j \in [I_z]$ *has an unknown distribution* $\mathbb{P}_j$ *partially characterized by an ambiguity set* $\mathcal{F}_j$, i.e., $\tilde{z}_j \sim \mathbb{P}_j \in \mathcal{F}_j \subseteq \mathcal{P}_0([\underline{z}_j, \bar{z}_j])$, $\underline{z}_j < \bar{z}_j$. *The distribution of* $\tilde{\boldsymbol{z}}$ *is the product distribution* $\mathbb{P} \in \mathcal{F}$, *where* $\mathcal{F} \triangleq \times_{j\in[I_z]} \mathcal{F}_j$. *We denote* $\mathcal{Z} = [\underline{\boldsymbol{z}}, \bar{\boldsymbol{z}}]$ *as the support set of all distributions in* $\mathcal{F}$, i.e., $\mathcal{F} \subseteq \mathcal{P}_0(\mathcal{Z})$. *We also partition the index set* $[I_z] = \mathcal{J}^+ \cup \mathcal{J}^- \cup \mathcal{J}$ *so that* $j \in \mathcal{J}^+$ *if and only if* $\underline{z}_j \geq 0$ *(*i.e., $\mathbb{P}[\tilde{z}_j \geq 0]$*),* $j \in \mathcal{J}^-$ *if and only if* $\bar{z}_j \leq 0$ *(*i.e., $\mathbb{P}[\tilde{z}_j \leq 0]$*).*

It is worth noting that the independence assumption may not be as restrictive as it initially seems. Under independent factors, we can model correlated uncertainty using a factor-based model, which is a common approach in operations research. For instance, in a portfolio optimization model, uncertain returns can be generated using a factor-based model such as $\tilde{\boldsymbol{r}} \triangleq \boldsymbol{F}\tilde{\boldsymbol{z}} + \boldsymbol{g}$, where $\boldsymbol{F}$ and $\boldsymbol{g}$ are fixed coefficients obtained from statistical methods or prior knowledge (see, *e.g.,* Natarajan et al. 2008). Even when the factor $\tilde{\boldsymbol{z}}$ has independent components, the uncertain returns can still exhibit strong correlation. The assumption of independently distributed random variables is prevalent in dynamic and stochastic optimization problems, and has been used in previous studies to justify the approximation of chance constrained problems using robust optimization methods (see Ben-Tal et al. 2009).

CARA certainty equivalent

Observe that under Assumption 1, the payoff function, $f(\boldsymbol{x}, \tilde{\boldsymbol{z}})$ is a random variable with ambiguous probability distributions. In considering both risk and ambiguity, we adopt the preference relation in Gilboa and Schmeidler (1989), such that for a given increasing utility function $u : \mathbb{R} \to \mathbb{R}$ and ambiguity set, $\mathcal{F} \subseteq \mathcal{P}_0(\mathcal{Z})$, the random payoff $f(\boldsymbol{x}_1, \tilde{\boldsymbol{z}})$ is preferred over $f(\boldsymbol{x}_2, \tilde{\boldsymbol{z}})$ if and only if the worst-case expected utility value $\inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[u(f(\boldsymbol{x}_1, \tilde{\boldsymbol{z}}))] \geq \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[u(f(\boldsymbol{x}_2, \tilde{\boldsymbol{z}}))]$. The goal of our robust decision model is to maximize the worst-case expected utility of the payoff as follows:

$$\max_{\boldsymbol{x}\in\mathcal{X}} \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}[u(f(\boldsymbol{x}, \tilde{\boldsymbol{z}}))]. \tag{1}$$

Specific to the CARA preferences, we have the exponential utility $u(v) = 1 - e^{-v/\kappa}$ with $\kappa > 0$ being the risk tolerance level. Since $u^{-1}(\cdot)$ is also increasing, the model (1) shares the same optimal solution with the model

$$\max_{\boldsymbol{x}\in\mathcal{X}} \inf_{\mathbb{P}\in\mathcal{F}} u^{-1}\left(\mathbb{E}_{\mathbb{P}}[u(f(\boldsymbol{x}, \tilde{\boldsymbol{z}}))]\right) = \max_{\boldsymbol{x}\in\mathcal{X}} \inf_{\mathbb{P}\in\mathcal{F}} -\kappa \log \mathbb{E}_{\mathbb{P}}\left[\exp\left(-\frac{f(\boldsymbol{x}, \tilde{\boldsymbol{z}})}{\kappa}\right)\right] \tag{2}$$

where the operator $u^{-1}(\mathbb{E}_{\mathbb{P}}[u(\cdot)])$ is referred to the *certainty equivalent* of the random payoffs under the utility function u and distribution $\mathbb{P} \in \mathcal{F}$. In articulating risk, certainty equivalent has the benefit of being more interpretable than expected utility. Hence, we focus on the robust CARA optimization model (2).

Definition 1. For a given random variable $\tilde{v}$, $\tilde{v} \sim \mathbb{P}$, the *CARA certainty equivalent* is defined as

$$\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}] \triangleq \begin{cases} \operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}] & \text{if } \kappa = 0 \\ \mathbb{E}_{\mathbb{P}}[\tilde{v}] & \text{if } \kappa = \infty \\ -\kappa \log \mathbb{E}_{\mathbb{P}}\left[\exp\left(-\dfrac{\tilde{v}}{\kappa}\right)\right] & \text{if } \kappa \in (0, \infty). \end{cases}$$

Note that the certainty equivalent would not exceed the risk-neutral expected payoffs as a result of Jensen's inequality for exponential utility. Given an ambiguity set of probability distributions $\mathcal{F}$, the *ambiguity-averse CARA certainty equivalent* is defined as

$$\mathbb{C}_{\mathcal{F}}^{\kappa}[\tilde{v}] \triangleq \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}].$$

Hence, Problem (2) can be written as

$$\max_{\boldsymbol{x} \in \mathcal{X}} \mathbb{C}_{\mathcal{F}}^{\kappa}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})]. \tag{3}$$

Throughout the paper, we focus on payoff functions that are finite for any $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{z} \in \mathcal{Z}$ so that $\mathbb{C}_{\mathcal{F}}^{\kappa}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})]$ is well-defined. Note that we extend the definition of CARA certainty equivalent to the cases of zero and infinite risk tolerance level, which would correspond to the worst-scenario and risk-neutral preferences, respectively. Specifically, our robust CARA optimization model recovers the distributionally robust optimization model of the form

$$\max_{\boldsymbol{x} \in \mathcal{X}} \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \tag{4}$$

when $\kappa = \infty$ where the decision maker is risk-neutral and ambiguity-averse. When $\kappa \in \mathbb{R}_+$, the robust CARA optimization model captures the preference of risk and ambiguity aversion, and the degree of risk aversion increases as the risk tolerance level κ decreases. In the extreme case of $\kappa = 0$, the model would coincide with the classical stochastic-free robust optimization model as follows:

$$\max_{\boldsymbol{x} \in \mathcal{X}} \inf_{\boldsymbol{z} \in \mathcal{Z}} f(\boldsymbol{x}, \boldsymbol{z}). \tag{5}$$

We note that as it is more common in the optimization literature to consider minimizing a cost function, we can also consider the following minmax certainty equivalent problem associated with exponential disutility:

$$\min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\mathbb{P} \in \mathcal{F}} \kappa \log \mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{f(\boldsymbol{x}, \tilde{\boldsymbol{z}})}{\kappa}\right)\right], \tag{6}$$

where the exponential disutility has the form $u(v) = e^{v/\kappa} - 1$. Hence, analogous to the definitions of CARA certainty equivalent and ambiguity-averse CARA certainty equivalent, we also define

$$\overline{\mathbb{C}}_{\mathbb{P}}^{\kappa}[\tilde{v}] \triangleq -\mathbb{C}_{\mathbb{P}}^{\kappa}[-\tilde{v}], \qquad \overline{\mathbb{C}}_{\mathcal{F}}^{\kappa}[\tilde{v}] \triangleq \sup_{\mathbb{P} \in \mathcal{F}} \overline{\mathbb{C}}_{\mathbb{P}}^{\kappa}[\tilde{v}] = -\mathbb{C}_{\mathcal{F}}^{\kappa}[-\tilde{v}].$$

In the remaining of the paper, we will adopt the payoff maximization decision model of Problem (3), although all the following development can easily be transformed to the cost minimization decision model of Problem (6).

We review some useful properties of the CARA certainty equivalent as follows:

PROPOSITION 1. *Consider the random variable* $\tilde{v}, \tilde{v} \sim \mathbb{P}$. *The CARA certainty equivalent has the following properties:*

1. $\lim_{\kappa \to \infty} \mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}] = \mathbb{E}_{\mathbb{P}}[\tilde{v}]$.
2. $\lim_{\kappa \downarrow 0} \mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}] = \operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}]$.
3. $\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}]$ *is non-decreasing in* $\kappa > 0$.
4. $\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}]$ *is jointly concave in* $\tilde{v}$ *and* $\kappa > 0$.
5. *For all* $\nu \in \mathbb{R}$, $\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v} + \nu] = \mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}] + \nu$.

These properties hold for $\mathbb{C}_{\mathcal{F}}^{\kappa}[\cdot]$ *as well.*

REMARK 1. The first two properties justify the definition of the CARA certainty equivalent at its limits. The third property relates to the monotonicity of the CARA certainty equivalent with regards to the degree of risk aversion. As we will show, we can exploit the concavity property for computational tractability. For a fix risk tolerance, the concavity of the CARA certainty equivalent implies a preference for diversification, which is also associated with risk aversion behavior. The last property implies *translation invariance*, *i.e.*, adding a constant amount to the random payoff would increase the CARA certainty equivalent by exactly the same amount. It is worth noting that the negative of the (ambiguity-averse) CARA certainty equivalent constitutes a *convex risk measure* (see, *e.g.*, Föllmer and Schied 2002), which is unique for CARA among all certainty equivalents.

We also present the following properties, which are useful for deriving exact and tractable approximations of robust CARA optimization models.

PROPOSITION 2. *Consider the random variables* $\tilde{v}, \tilde{\nu}$, $(\tilde{v}, \tilde{\nu}) \sim \mathbb{P}$.

1. $\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}]$ *is super-additive in* $(\kappa, \tilde{v})$, i.e.,

$$\mathbb{C}_{\mathbb{P}}^{\kappa_1 + \kappa_2}[\tilde{v} + \tilde{\nu}] \geq \mathbb{C}_{\mathbb{P}}^{\kappa_1}[\tilde{v}] + \mathbb{C}_{\mathbb{P}}^{\kappa_2}[\tilde{\nu}]$$

   for any $\kappa_1, \kappa_2 \in \mathbb{R}_+$.
2. *If* $\tilde{v}$, $\tilde{\nu}$ *are also independently distributed, then*

$$\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v} + \tilde{\nu}] = \mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}] + \mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{\nu}].$$

These properties hold for $\mathbb{C}_{\mathcal{F}}^{\kappa}[\cdot]$ *as well.*

Payoff functions with affine perturbations

We now focus on robust CARA optimization models with affinely perturbed payoff functions, which arise, for example, in shortest path (Chicoisne et al. 2018) and portfolio optimization (El Ghaoui et al. 2003). Specifically, we consider the payoff function of the following form,

$$f(\boldsymbol{x},\boldsymbol{z}) = a^0(\boldsymbol{x}) + \sum_{j\in[I_z]} a^j(\boldsymbol{x}) z_j. \tag{7}$$

We assume the function $a^j(\boldsymbol{x}) : \mathbb{R}^{I_x} \to \mathbb{R}$ is concave for all $j \in \{0\}\cup\mathcal{J}^+$, convex for all $j\in\mathcal{J}^-$, and affine for all $j\in\mathcal{J}$ so that $f(\boldsymbol{x},\boldsymbol{z})$ is concave in $\boldsymbol{x}$ for any $\boldsymbol{z}\in\mathcal{Z}$.

Since $\tilde{z}_j$'s are independent, we have $\mathbb{C}^\kappa_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = a^0(\boldsymbol{x}) + \sum_{j\in[I_z]}\phi_j(\kappa, a^j(\boldsymbol{x}))$ where we define the function $\phi_j : [0,\infty]\times\mathbb{R}\to\mathbb{R}$,

$$\phi_j(\kappa,\lambda) \triangleq \mathbb{C}^\kappa_{\mathcal{F}_j}[\lambda\tilde{z}_j]. \tag{8}$$

By introducing the auxiliary variables λ^0 and λ^j, $j\in[I_z]$, we obtain the following representation

$$\begin{aligned}
\mathbb{C}^\kappa_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \max_{\boldsymbol{\lambda}\in\mathbb{R}^{1+I_z}} \quad & \lambda^0 + \sum_{j\in[I_z]}\phi_j(\kappa,\lambda^j) \\
\text{s.t.} \quad & a^j(\boldsymbol{x}) \geq \lambda^j \qquad \forall j\in\{0\}\cup\mathcal{J}^+ \\
& a^j(\boldsymbol{x}) \leq \lambda^j \qquad \forall j\in\mathcal{J}^- \\
& a^j(\boldsymbol{x}) = \lambda^j \qquad \forall j\in\mathcal{J},
\end{aligned} \tag{9}$$

As we see, having stochastic independent factors would help us decompose the multi-dimensional integration problems associated with evaluating the CARA certainty equivalent to more analytically tractable single-dimensional integration problems. To appreciate the simplification, we consider the following example.

Example 1. Let $\tilde{z}_j$ be independent uniform random variables over the unit interval $[0,1]$, $a^j(\boldsymbol{x}) \equiv a_j < 0$ for any $j\in[I_z]$ and $a^0(\boldsymbol{x}) \equiv a_0 > 0$, then

$$\mathbb{C}^\kappa_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = a_0 - \sum_{j\in[I_z]}\kappa\log\int_0^1 \exp\left(-\frac{a_j z_j}{\kappa}\right)\mathrm{d}z_j = a_0 - \sum_{j\in[I_z]}\kappa\log\left(\frac{\kappa-\kappa e^{-a_j/\kappa}}{a_j}\right).$$

Hence the computation of $\mathbb{C}^\kappa_{\mathbb{P}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})]$ can be obtained in closed form. In contrast, due to multi-dimensional integration, evaluating an expected concave piecewise linear utility such as

$$\mathbb{E}_{\mathbb{P}}\left[\min\left\{a_0 + \sum_{j\in[I_z]} a_j\tilde{z}_j, 0\right\}\right]$$

is known to be a #P-hard problem (Dyer and Stougie 2006, Hanasusanto et al. 2016).

Finally, we point out the unique computational challenge of robust CARA optimization compared with distributionally robust optimization. We consider an affine payoff function $f(\boldsymbol{x},\boldsymbol{z}) = \boldsymbol{x}^\top \boldsymbol{z}$ under an ambiguity set with mean and polytope support set, *i.e.*,

$$\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}^{I_z}) \;\middle|\; \begin{array}{l} \tilde{\boldsymbol{z}} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{\boldsymbol{z}}] = \boldsymbol{\mu} \\ \mathbb{P}[\tilde{\boldsymbol{z}} \in \mathcal{Z}] = 1 \end{array} \right\}. \tag{10}$$

When $\kappa = 0$ or $\kappa = \infty$, the ambiguity-averse CARA certainty equivalent $\mathbb{C}_{\mathcal{G}}^{\kappa}[\boldsymbol{x}^\top \tilde{\boldsymbol{z}}]$ can be easily evaluated computationally, corresponding to $\inf_{\boldsymbol{z}\in\mathcal{Z}} \boldsymbol{x}^\top \boldsymbol{z}$ and $\boldsymbol{x}^\top \boldsymbol{\mu}$, respectively. For any $\boldsymbol{x} \in \mathcal{X}$, and $\kappa \in (0,\infty)$, we have

$$\begin{aligned}
& \mathbb{C}_{\mathcal{G}}^{\kappa}[\boldsymbol{x}^\top \tilde{\boldsymbol{z}}] \geq t \\
\iff & -\kappa \log \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[\exp\left(-\frac{\boldsymbol{x}^\top \tilde{\boldsymbol{z}}}{\kappa}\right)\right] \geq t \\
\iff & \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[\kappa \exp\left(\frac{t - \boldsymbol{x}^\top \tilde{\boldsymbol{z}}}{\kappa}\right)\right] \leq \kappa \\
\iff & \exists\ \alpha, \boldsymbol{\beta} \text{ such that } \begin{cases} \alpha + \boldsymbol{\beta}^\top \boldsymbol{\mu} \leq \kappa \\ \alpha + \boldsymbol{\beta}^\top \boldsymbol{z} \geq \kappa \exp\left(\frac{t - \boldsymbol{x}^\top \boldsymbol{z}}{\kappa}\right) \ \forall \boldsymbol{z} \in \mathcal{Z}. \end{cases}
\end{aligned}$$

where the last step follows from duality. Note that the constraint

$$\max_{\boldsymbol{z}\in\mathcal{Z}} \kappa \exp\left(\frac{t - \boldsymbol{x}^\top \boldsymbol{z}}{\kappa}\right) - \alpha - \boldsymbol{\beta}^\top \boldsymbol{z} \leq 0 \tag{11}$$

involves a convex maximization problem, which is in general an intractable optimization problem.

THEOREM 1. *Evaluating the ambiguity-averse CARA certainty equivalent of an affine payoff function over the polyhedral mean-support ambiguity set* (10) *is NP-hard for some* $\kappa \in (0,\infty)$.

To derive a tractable convex approximation for the ambiguity-averse CARA certainty equivalent, we can consider the reformulation-perspectification approach of Bertsimas et al. (2022). We can also consider a hybrid uncertainty model, where some random variables are independently distributed, and use infimum convolution to obtain a unified bound on the ambiguity-averse CARA certainty equivalent (see Wiesemann et al. 2014, Goh and Sim 2010). However, these approaches would inevitably increase the complexity of the model. In this paper, we focus primarily on independently distributed random variables as a foundation for constructing more complex models.

Practicably solvable reformulations

We next focus on tractable reformulations of robust CARA optimization problems under Assumption 1. From a theoretical perspective, an optimization problem is considered to be tractable if it can be solved in polynomial time (*e.g.*, by the ellipsoid method). However, this definition does not always reflect how well optimization problems can be solved in practice. Therefore, we use the term *practicably solvable* problems informally to refer to problems that are of practical interest and can be efficiently solved by current off-the-shelf solvers. Our main interest is to obtain solutions to the robust CARA optimization problems reliably and within reasonable time so that the solutions can be implemented in practice. In this regard, we refer to a problem as practicably solvable if it can be formulated using a modest number of decision variables, and a modest number of linear, convex quadratic, second-order conic, exponential conic constraints that can be solved using current off-the-shelf software tools.[1].Among these constraint types, exponential conic constraints are essential in modeling exponential and logarithms arising from the robust CARA optimization model. Exponential conic constraints are now supported in Mosek, and they can also be approximated fairly accurately via second-order conic constraints (see, *e.g.*, Ye and Xie 2021), which are broadly supported in solvers such as Gurobi, CPLEX and SDPT3.

Definition 2. A convex set $\mathcal{W} \subseteq \mathbb{R}^I$ is *exponential cone representable ($\mathcal{K}_{\exp}$-representable)* if it is conic representable with exponential cones, *i.e.*,

$$\boldsymbol{x} \in \mathcal{W} \Longleftrightarrow \exists\ \boldsymbol{u} \in \mathbb{R}^J : \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{b} \in \mathcal{K}_{\exp}^K$$

with $\boldsymbol{A} \in \mathbb{R}^{3K\times I}$, $\boldsymbol{B} \in \mathbb{R}^{3K\times J}$, $\boldsymbol{b} \in \mathbb{R}^{3K}$ and $\mathcal{K}_{\exp}^K$ is the Cartesian product of K *exponential cones*

$$\mathcal{K}_{\exp} \triangleq \{(x_1, x_2, x_3) | x_1 \geq x_2 \exp(x_3/x_2), x_2 > 0\} \cup \{(x_1, 0, x_3) | x_1 \geq 0, x_3 \leq 0\}. \tag{12}$$

We also say that a convex (concave) function is $\mathcal{K}_{\exp}$-representable if its epigraph (hypograph) is a $\mathcal{K}_{\exp}$-representable set.

Theorem 2. *Let $g(\boldsymbol{x}, \kappa) = -\kappa \log \sum_{i\in[I]} p_i e^{-x_i/\kappa}$ with $\kappa > 0$ and $p_i > 0$ for all $i \in [I]$, then the closure of its hypograph $\{(\boldsymbol{x}, \kappa, y) : y \leq g(\boldsymbol{x}, \kappa), \kappa > 0\}$ can be represented by*

$$\left\{ (\boldsymbol{x}, \kappa, y) \ \middle|\ \exists\ \boldsymbol{q} \in \mathbb{R}^I \ : \ \sum_{i\in[I]} p_i q_i \leq \kappa,\ (q_i, \kappa, y - x_i) \in \mathcal{K}_{\exp} \quad \forall i \in [I] \right\}.$$

For Problem (9) to be practicably solvable, we note that the functions $\phi_j(\kappa, \lambda^j)$, $j \in [I_z]$ are $\mathcal{K}_{\exp}$-representable for some choice of ambiguity sets, as we will show in the following example.

Example 2. Consider an ambiguity set with mean, mean absolute deviation, and support information

$$\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathbb{R}) \ \middle|\ \begin{array}{l} \tilde{z} \sim \mathbb{P} \\ \mathbb{E}_{\mathbb{P}}[\tilde{z}] = \mu \\ \mathbb{E}_{\mathbb{P}}[|\tilde{z} - \mu|] \leq \delta \\ \mathbb{P}[\tilde{z} \in [-1, 1]] = 1 \end{array} \right\}$$

[1] We do not include semidefinite constraints since semidefinite programs (SDPs) are less scalable compared to aforementioned problems (see, *e.g.,* Toh 2018, for a rough guide of the SDP problem size that can be solved by current off-the-shelf solvers)

where the support is normalized without loss of generality, otherwise we can consider transformation $z \mapsto \frac{2z-(\underline{z}+\overline{z})}{\overline{z}-\underline{z}}$ if $\mathbb{P}[\tilde{z} \in [\underline{z}, \overline{z}]] = 1$. As in Postek et al. (2018) where the inequality in $\mathcal{G}$ is replaced by equality, we show

$$\phi(\kappa, \lambda) = -\kappa \log\left(\frac{\delta}{2(\mu+1)} e^{\lambda/\kappa} + \frac{\delta}{2(1-\mu)} e^{-\lambda/\kappa} + \left(1 - \frac{\delta}{2(\mu+1)} - \frac{\delta}{2(1-\mu)}\right) e^{-\mu\lambda/\kappa}\right),$$

where

$$\phi(\kappa, \lambda) \triangleq \mathbb{C}_{\mathcal{G}}^{\kappa}[\lambda \tilde{z}]$$

is hence $\mathcal{K}_{\exp}$-representable based on the Theorem 2. The proof is relegated to Appendix A.

We refer interested readers to Nemirovski and Shapiro (2007) for more examples, which we also summarize in Table 1 and discuss their exponential conic representations in Appendix B.

Payoff functions with concave piecewise affine perturbations

We have shown that robust CARA optimization is practicably solvable for concave payoff functions with affine perturbations. However, in practice, payoff functions can be nonlinear in the uncertain factors. Many real-world problems, such as option pricing (Bertsimas and Popescu 2002), inventory management (Ardestani-Jaafari and Delage 2016), routing optimization with time windows (Zhang et al. 2021), and electric vehicle charging scheduling (Chen et al. 2023), involve payoff functions with concave piecewise affine perturbations. Piecewise affine functions are also fundamental building blocks in stochastic and dynamic optimization models commonly used in operations research. For instance, a two-stage linear optimization problem with fixed recourse can have its recourse function represented by a piecewise affine function with exponentially many pieces (see, *e.g.*, Birge and Louveaux 2011).

Therefore, in contrast to existing literature that focuses on affine perturbations (see, *e.g.*, Nemirovski and Shapiro 2007, Jaillet et al. 2016), we consider payoff functions with concave piecewise affine perturbations, *i.e.*,

$$f(\boldsymbol{x}, \boldsymbol{z}) = \min_{i \in \mathcal{I}} \left\{ a_i^0(\boldsymbol{x}) + \sum_{j \in [I_z]} a_i^j(\boldsymbol{x}) z_j \right\}. \tag{13}$$

For each $i \in \mathcal{I}$, we assume the function $a_i^j(\boldsymbol{x}) : \mathbb{R}^{I_x} \to \mathbb{R}$ is concave for all $j \in \{0\} \cup \mathcal{J}^+$, convex for all $j \in \mathcal{J}^-$, and affine for all $j \in \mathcal{J}$. Moreover, we assume the optimization problem $\max_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}, \boldsymbol{z})$ is practicably solvable for any $\boldsymbol{z} \in \mathcal{Z}$.

However, when dealing with payoff functions that have concave piecewise affine perturbations, the robust CARA optimization model becomes intractable. Example 1 demonstrates that even evaluating the CARA certainty equivalent $\mathbb{C}_{\mathbb{P}}^{\infty}\left[\min\left\{a_0 + \sum_{j \in [I_z]} a_j \tilde{z}_j, 0\right\}\right]$ under independent uniform distribution can be #P-hard. As a result, we are motivated to develop a practicably solvable lower bound for the ambiguity-averse CARA certainty equivalent of the payoff function (13).

Theorem 3. *The ambiguity-averse CARA certainty equivalent of the payoff function* (13), $\mathbb{C}^{\kappa}_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})]$ *has a practicably solvable lower bound:*

$$\begin{aligned}\Lambda(\kappa,\boldsymbol{x}) \triangleq \max_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ & \Phi(\kappa,\boldsymbol{\alpha},\boldsymbol{\beta}) \\ \text{s.t.} \ & a_i^0(\boldsymbol{x}) \geq \alpha_i & \forall i \in \mathcal{I} \\ & a_i^j(\boldsymbol{x}) \geq \beta_i^j & \forall i \in \mathcal{I}, j \in \mathcal{J}^+ \\ & a_i^j(\boldsymbol{x}) \leq \beta_i^j & \forall i \in \mathcal{I}, j \in \mathcal{J}^- \\ & a_i^j(\boldsymbol{x}) = \beta_i^j & \forall i \in \mathcal{I}, j \in \mathcal{J} \\ & \boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{I}|}, \boldsymbol{\beta} \in \mathbb{R}^{|\mathcal{I}| \times I_z}, \end{aligned} \tag{14}$$

where the objective function is

$$\begin{aligned}\Phi(\kappa,\boldsymbol{\alpha},\boldsymbol{\beta}) \triangleq \max_{\rho,\boldsymbol{\gamma},\boldsymbol{r},\boldsymbol{q},\boldsymbol{\kappa}} \ & r_0 + \rho \\ \text{s.t.} \ & \kappa_0 + \kappa_1 = \kappa \\ & \sum_{i \in \mathcal{I}} q_i \leq \kappa_1 \\ & (q_i, \kappa_1, \rho - r_i) \in \mathcal{K}_{\exp} & \forall i \in \mathcal{I} \\ & \sum_{j \in [I_z]} \phi_j(\kappa_0, \gamma^j) \geq r_0 \\ & \alpha_i + \sum_{j \in [I_z]} \phi_j(\kappa_1, \beta_i^j - \gamma^j) \geq r_i & \forall i \in \mathcal{I} \\ & \boldsymbol{\gamma} \in \mathbb{R}^{I_z}, \boldsymbol{r} \in \mathbb{R}^{1+|\mathcal{I}|}, \boldsymbol{\kappa} \in \mathbb{R}^2_+, \rho \in \mathbb{R}, \boldsymbol{q} \in \mathbb{R}^{|\mathcal{I}|}. \end{aligned} \tag{15}$$

It is worth noting that despite being challenging to determine the precise value of $\mathbb{C}^{\kappa}_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})]$, the following result provides conditions when the the bound in Theorem 3 can be exact. We also provide two examples in Appendix C to demonstrate its advantages by comparing with other approximations including a bound in Nemirovski and Shapiro (2007) and the Monte-Carlo approximations.

Theorem 4. *For any* $\boldsymbol{x} \in \mathcal{X}$*, the function* $\Lambda(\kappa,\boldsymbol{x})$ *is non-decreasing in* $\kappa \in [0,\infty]$ *and satisfies* $\Lambda(\kappa,\boldsymbol{x}) \geq \inf_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{x},\boldsymbol{z})$*. Moreover,* $\mathbb{C}^{\kappa}_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \Lambda(\kappa,\boldsymbol{x})$ *if there exists some* $i^* \in \mathcal{I}$ *such that*

$$a_{i^*}^0(\boldsymbol{x}) + \sum_{j\in[I_z]} a_{i^*}^j(\boldsymbol{x}) z_j \leq a_i^0(\boldsymbol{x}) + \sum_{j\in[I_z]} a_i^j(\boldsymbol{x}) z_j \qquad \forall \boldsymbol{z} \in \mathcal{Z}, i \in \mathcal{I}.$$

Remark 2. In the extreme risk aversion where $\kappa = 0$, Theorem 4 implies that

$$\inf_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{x},\boldsymbol{z}) = \mathbb{C}^{0}_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = \Lambda(0,\boldsymbol{x}) \geq \inf_{\boldsymbol{z}\in\mathcal{Z}} f(\boldsymbol{x},\boldsymbol{z}),$$

alluding to the improving accuracy of the approximation $\Lambda(\kappa,\boldsymbol{x})$ as the risk tolerance, κ decreases. The exact result could also occur when there exists a dominant payoff component, $i^* \in \mathcal{I}$ as defined in Theorem 4, which could arise in situations with low coefficient of variations among the payoff components, or when there are identical variations such that $a_{i_1}^j(\boldsymbol{x}) = a_{i_2}^j(\boldsymbol{x})$ for all $i_1, i_2 \in \mathcal{I}, j \in [I_z]$.

3. Adaptive optimization and tractable approximations

We now propose a more general framework for adaptive robust CARA optimization that has provisions for recourse. We focus on a two-stage adaptive optimization problem with payoff function defined by the optimal value of a linear optimization problem as follows,

$$\begin{aligned} f(\boldsymbol{x},\boldsymbol{z}) = \max_{\boldsymbol{y}} \ & \boldsymbol{c}^\top \boldsymbol{y} \\ \text{s.t.} \ & \boldsymbol{b}_i^\top \boldsymbol{y} \le a_i^0(\boldsymbol{x}) + \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z} \ \forall i \in \mathcal{I}, \\ & \boldsymbol{y} \in \mathbb{R}^{I_y} \end{aligned} \tag{16}$$

where $\mathcal{I}$ is the index set of constraints and $\boldsymbol{a}_i(\boldsymbol{x})$ is the vector of $a_i^j(\boldsymbol{x})$ of $j \in [I_z]$ for each $i \in \mathcal{I}$. In this problem, $\boldsymbol{x}$ is the *here-and-now* decision and $\boldsymbol{y}$ is the *wait-and-see* or *recourse* decision adapted to uncertain parameter $\tilde{\boldsymbol{z}}$. As before, for each $i \in \mathcal{I}$, we assume the function $a_i^j(\boldsymbol{x}) : \mathbb{R}^{I_x} \to \mathbb{R}$ is concave for all $j \in \{0\} \cup \mathcal{J}^+$, convex for all $j \in \mathcal{J}^-$, and affine for all $j \in \mathcal{J}$.

Observe that with $I_y = 1$, $\boldsymbol{c} = 1$, and $\boldsymbol{b}_i = 1$, $i \in \mathcal{I}$, the concave piecewise affine payoff function (13) is a special case of (16). Moreover, we can assume without any loss of generality that the objective function in Problem (16) contains only the recourse decision. Otherwise, for the objective function $\boldsymbol{c}^\top \boldsymbol{y} + a_0^0(\boldsymbol{x}) + \boldsymbol{a}_0^\top(\boldsymbol{x})\boldsymbol{z}$, we can introduce another auxiliary variable y_{I_y+1} to replace it, and add the constraint $y_{I_y+1} - \boldsymbol{c}^\top \boldsymbol{y} \le a_0^0(\boldsymbol{x}) + \boldsymbol{a}_0^\top(\boldsymbol{x})\boldsymbol{z}$. We assume without loss of generality that $\boldsymbol{b}_i \neq \boldsymbol{0}$, $i \in \mathcal{I}$; otherwise such a constraint can always be incorporated in $\mathcal{X}$, which describes the feasible set of the here-and-now decision.

Drawing from the insights of Zhen et al. (2018), we can always improve the formulation of an adaptive optimization problem via *Fourier-Motzkin elimination* of the recourse variables whenever it is computationally viable to do so. In particular, we highlight that Problem (16) does not have any equality constraint because for each equality constraint, we can eliminate a recourse variable without increasing the size of the formulation. The two-stage optimization problem is general enough to cover many practical optimization problems such as appointment scheduling, network lot-sizing, projection management, and so forth.

Definition 3. We say Problem (16) has *complete recourse* if and only if for any $\boldsymbol{d} \in \mathbb{R}^{|\mathcal{I}|}$, there exists some $\boldsymbol{y} \in \mathbb{R}^{I_y}$ such that $\boldsymbol{b}_i^\top \boldsymbol{y} \le d_i$ for all $i \in \mathcal{I}$ (see, *e.g.,* Birge and Louveaux 2011).

Assumption 2. *We assume that the optimization problem* $\max_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x},\boldsymbol{z})$ *is practicably solvable and is bounded from above for any* $\boldsymbol{z} \in \mathcal{Z}$.

Note that we can express the ambiguity-averse CARA certainty equivalent as the following optimization problem,

$$\begin{aligned} \mathbb{C}_{\mathcal{F}}^{\kappa}\left[f(\boldsymbol{x},\tilde{\boldsymbol{z}})\right] = \max_{\boldsymbol{y}} \ & \mathbb{C}_{\mathcal{F}}^{\kappa}\left[\boldsymbol{c}^\top \boldsymbol{y}(\tilde{\boldsymbol{z}})\right] \\ \text{s.t.} \ & \boldsymbol{b}_i^\top \boldsymbol{y}(\boldsymbol{z}) \le a_i^0(\boldsymbol{x}) + \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z} \quad \forall \boldsymbol{z} \in \mathcal{Z}, \quad \forall i \in \mathcal{I} \\ & \boldsymbol{y} \in \mathcal{R}^{I_z, I_y}. \end{aligned} \tag{17}$$

Since $\boldsymbol{y}$ is a function map instead of a finite vector of decision variables, the above problem is generally intractable.

A common approach to solve the adaptive optimization problem approximately is to use *linear decision rule (LDR)* to restrict the recourse function map to be affinely dependent on the uncertain parameters, *i.e.*, $\boldsymbol{y} \in \mathcal{L}^{I_z, I_y}$ where

$$\mathcal{L}^{I_z, I_y} \triangleq \left\{ \boldsymbol{y} \in \mathcal{R}^{I_z, I_y} \;\middle|\; \exists\, \boldsymbol{y}^0, \boldsymbol{y}^1, \ldots, \boldsymbol{y}^{I_z} : \boldsymbol{y}(\boldsymbol{z}) = \boldsymbol{y}^0 + \sum_{j \in [I_z]} z_j \boldsymbol{y}^j \right\}. \tag{18}$$

However, it has been well known that such approximation can be rather conservative and in some situations, we may sacrifice too much for tractability (Garstka and Wets 1974). While there is a long list of tractable decision rules that have been proposed to improve upon LDRs in the literature, the main challenge lies in applying these rules to the robust CARA optimization framework, which is inherently a nonlinear problem. Many decision rules are unique to classical robust optimization and are incompatible with distribution ambiguity, such as the extended decision rule of Chen and Zhang (2009) and the piecewise linear decision rule of Ben-Tal et al. (2020). Additionally, it is not clear how to apply the lifted decision rules of Georghiou et al. (2015), Wiesemann et al. (2014) and the segregated decision rules of Chen et al. (2008) in their generality under CARA preferences.

Fortunately, the deflected linear decision rule (DLDR) (Chen et al. 2008, Goh and Sim 2010) can be applied to Problem (17) and has shown to be effective in improving upon LDRs. For this purpose, we solve for each $i \in \mathcal{I}$:

$$\begin{aligned} \max_{\boldsymbol{y} \in \mathbb{R}^{I_y}} \;& \boldsymbol{c}^\top \boldsymbol{y} \\ & \boldsymbol{b}_k^\top \boldsymbol{y} \le 0 \quad \forall k \in \mathcal{I} \backslash \{i\} \\ & \boldsymbol{b}_i^\top \boldsymbol{y} = -1 \end{aligned} \tag{19}$$

and define $\mathcal{I}^o \subseteq \mathcal{I}$ as the index set of i such that the above optimization problem is feasible and $\boldsymbol{y}_\diamond^i$ as the corresponding optimal solution. In the case of complete recourse, Chen et al. (2008) note that Problem (19) would always be feasible. The DLDR has the following form:

$$\boldsymbol{y}^\dagger(\boldsymbol{z}) \triangleq \bar{\boldsymbol{y}}(\boldsymbol{z}) + \sum_{i \in \mathcal{I}^o} \boldsymbol{y}_\diamond^i \left(h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z}) \right)^+ \tag{20}$$

where

$$\begin{aligned} \bar{\boldsymbol{y}}(\boldsymbol{z}) &\triangleq \boldsymbol{y}^0 + \boldsymbol{Y}\boldsymbol{z} \\ h_i(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}) &\triangleq \boldsymbol{b}_i^\top \boldsymbol{y} - a_i^0(\boldsymbol{x}) - \boldsymbol{a}_i^\top(\boldsymbol{x}) \boldsymbol{z} \;\; \forall i \in \mathcal{I}, \end{aligned}$$

with $\boldsymbol{y}^0 \in \mathbb{R}^{I_y}$ and $\boldsymbol{Y} \triangleq [\boldsymbol{y}^1, \ldots, \boldsymbol{y}^{I_z}] \in \mathbb{R}^{I_y \times I_z}$.

Multi-deflected linear decision rule

We now propose the *multi-deflected linear decision rule (MLDR)* that improves upon DLDR. Specifically, we first solve for each $i \in \mathcal{I}$:

$$\begin{aligned} \max_{\boldsymbol{y} \in \mathbb{R}^{I_y}} \ & \boldsymbol{c}^\top \boldsymbol{y} \\ & \boldsymbol{b}_k^\top \boldsymbol{y} \le 0 \qquad \forall k \in \mathcal{I} \backslash \{i\} \\ & \boldsymbol{b}_i^\top \boldsymbol{y} = -\|\boldsymbol{b}_i\| \end{aligned} \tag{21}$$

and denote $\boldsymbol{y}_*^i$ as its optimal solution for each $i \in \mathcal{I}^o$. Observe that $\boldsymbol{y}_*^i = \|\boldsymbol{b}_i\| \boldsymbol{y}_\diamond^i$. Then we partition the index set $\mathcal{I}^o$ as

$$\mathcal{I}^o = \bigcup_{\ell \in [m]} \mathcal{I}_\ell^o$$

such that $\boldsymbol{y}_*^{i_1} = \boldsymbol{y}_*^{i_2}$ if and only if i_1 and i_2 are in the same $\mathcal{I}_\ell^o$. We denote $\boldsymbol{y}_*^\ell$ as any $\boldsymbol{y}_*^i$ with $i \in \mathcal{I}_\ell^o$ and define MLDR as follows:

$$\hat{\boldsymbol{y}}(\boldsymbol{z}) \triangleq \bar{\boldsymbol{y}}(\boldsymbol{z}) + \sum_{\ell \in [m]} \boldsymbol{y}_*^\ell \left(\max_{i \in \mathcal{I}_\ell^o} \left\{ \frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|} \right\} \right)^+ . \tag{22}$$

In the following, we show how the MLDR can improve over the DLDR.

Theorem 5. *Under Assumption 2, for any distribution $\mathbb{P} \in \mathcal{P}_0(\mathcal{Z})$ and $\kappa \in \mathbb{R}_+$, we have $\boldsymbol{c}^\top \boldsymbol{y}_*^\ell \le 0$ for all $\ell \in [m]$ and*

$$\mathbb{C}_{\mathbb{P}}^{\kappa} \left[\boldsymbol{c}^\top \hat{\boldsymbol{y}}(\tilde{\boldsymbol{z}}) \right] \ge \mathbb{C}_{\mathbb{P}}^{\kappa} \left[\boldsymbol{c}^\top \boldsymbol{y}^\dagger(\tilde{\boldsymbol{z}}) \right].$$

We also show that the MLDR can replicate the optimal recourse function for the simplest class of adaptive optimization problems with complete recourse.

Theorem 6. *Suppose Problem* (16) *has complete recourse and $I_y = 1$, then there exists an MLDR that is optimal in Problem* (16) *for all $\boldsymbol{z} \in \mathcal{Z}$.*

Remark 3. We remark that for the same class of adaptive optimization problems with complete recourse, Bertsimas et al. (2019) has also proposed the lifted *affine recourse adaptation* (ARA), which can achieve the optimal worst-case risk-neutral objective value under a moment-based ambiguity set. However, unlike MLDR, the lifted ARA may not necessarily replicate the optimal recourse function that we have for MLDR in Theorem 6. Hence, Bertsimas et al. (2019) caution the use of lifted ARA as a form of decision rule or policy for multi-period decision making.

In line with Theorem 6, we illustrate the advantage of MLDR over DLDR in the following example.

Example 3. Consider the payoff function as follows:

$$\begin{aligned} f(x,z) = \max_{y} \ & y \\ \text{s.t.} \ & y \le 2 \\ & y \le z \\ & y \le 2z - 1 \\ & y \le 3z \end{aligned}$$

where the optimal decision rule is $y^{OPT}(z) = \min\{2, z, 2z-1, 3z\}$. Note that $y_*^i = -1$ for all $i \in \mathcal{I}^o = \mathcal{I}$. It is easy to see $y^{OPT}(z)$ can be expressed as an MLDR, such as

$$\hat{y}(z) = 2 - (\max\{0, 2-z, 3-2z, 2-3z\})^+.$$

However, it cannot be represented by any DLDR, which has the form

$$y^\dagger(z) = \bar{y}(z) - (\bar{y}(z) - 2)^+ - (\bar{y}(z) - z)^+ - (\bar{y}(z) - 2z + 1)^+ - (\bar{y}(z) - 3z)^+$$

for any $\bar{y}(z) = y^0 + y^1 z$. To see this, note that

$$y^\dagger(z) \le \min\{y^{OPT}(z), 5z - 1 - \bar{y}(z), 3z + 1 - 2\bar{y}(z)\}.$$

To guarantee $y^\dagger(z) = y^{OPT}(z)$ for all $z \in \mathbb{R}$, the slope of $5z - 1 - \bar{y}(z)$ and $3z + 1 - 2\bar{y}(z)$ must lie in the interval $[0,3]$, which implies $5 - y^1 \le 3$ and $3 - 2y^1 \ge 0$, a contradiction. Hence there always exists $z \in \mathbb{R}$ such that $y^\dagger(z) < y^{OPT}(z)$ under any choice of $y^0, y^1 \in \mathbb{R}$.

In the following proposition, we establish the conditions of feasibility of the MLDR in Problem (17).

Proposition 3. *Suppose* $\bar{\boldsymbol{y}} \in \mathcal{L}^{I_z, I_y}$ *satisfies*

$$\boldsymbol{b}_i^\top \bar{\boldsymbol{y}}(\boldsymbol{z}) \le a_i^0(\boldsymbol{x}) + \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z} \quad \forall \boldsymbol{z} \in \mathcal{Z}, \quad \forall i \in \mathcal{I} \backslash \mathcal{I}^o,$$

Then the MLDR, $\hat{\boldsymbol{y}}$ *satisfies*

$$\boldsymbol{b}_i^\top \hat{\boldsymbol{y}}(\boldsymbol{z}) \le a_i^0(\boldsymbol{x}) + \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z} \quad \forall \boldsymbol{z} \in \mathcal{Z}, \quad \forall i \in \mathcal{I}.$$

Therefore, by applying the MLDR to Problem (17), we obtain a lower bound of (17) as follows:

$$\begin{aligned} \mathbb{C}_{\mathcal{F}}^{\kappa}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})] \ge \max_{\bar{\boldsymbol{y}}} \ & \mathbb{C}_{\mathcal{F}}^{\kappa}\left[\boldsymbol{c}^\top\left(\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}) + \sum_{\ell \in [m]} \boldsymbol{y}_*^\ell \left(\max_{i \in \mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}), \tilde{\boldsymbol{z}})}{\|\boldsymbol{b}_i\|}\right\}\right)^+\right)\right] \\ \text{s.t.} \ & h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z}) \le 0 \qquad \forall \boldsymbol{z} \in \mathcal{Z}, \quad \forall i \in \mathcal{I} \backslash \mathcal{I}^o \\ & \bar{\boldsymbol{y}} \in \mathcal{L}^{I_z, I_y}. \end{aligned} \tag{23}$$

However, Problem (23) involves evaluation of ambiguity-averse CARA certainty equivalent of a sum of concave piecewise affine (*i.e.*, sum-of-min) functions. Theoretically, we can write the sum-of-min functions as concave piecewise affine functions so that the problem can be approximated using techniques in Theorem 3. However, this may not be practical since the piecewise affine reformulation might involve exponentially many pieces. Nevertheless, we provide a tractable lower bound as follows.

THEOREM 7. *Under Assumption 2, the ambiguity-averse CARA certainty equivalent* (17) *has a practicably solvable lower bound:*

$$
\begin{array}{lll}
\max\limits_{\substack{\boldsymbol{\kappa},\boldsymbol{r},\boldsymbol{\alpha},\boldsymbol{\beta}\\ \boldsymbol{y}^0,\boldsymbol{Y},\bar{\boldsymbol{\lambda}},\underline{\boldsymbol{\lambda}}}} & r_0 - \sum\limits_{\ell\in[m]}(\boldsymbol{c}^\top \boldsymbol{y}_*^\ell) r_\ell & \\
\text{s.t.} & \kappa_0 - \sum\limits_{\ell\in[m]}(\boldsymbol{c}^\top \boldsymbol{y}_*^\ell)\kappa_\ell = \kappa & \\
& \boldsymbol{c}^\top \boldsymbol{y}^0 + \sum\limits_{j\in[I_z]} \phi_j(\kappa_0, \boldsymbol{c}^\top \boldsymbol{y}^j) \geq r_0 & \\
& \Phi(\kappa_\ell,\ \bar{\boldsymbol{\alpha}}_{\mathcal{I}_\ell^o} - \bar{\boldsymbol{b}}_{\mathcal{I}_\ell^o}^\top \boldsymbol{y}^0,\ \bar{\boldsymbol{\beta}}_{\mathcal{I}_\ell^o} - \bar{\boldsymbol{b}}_{\mathcal{I}_\ell^o}^\top \boldsymbol{Y}) \geq r_\ell & \forall \ell\in[m] \\
& \bar{\boldsymbol{\lambda}}_i^\top \bar{\boldsymbol{z}} - \underline{\boldsymbol{\lambda}}_i^\top \underline{\boldsymbol{z}} \leq \alpha_i - \boldsymbol{b}_i^\top \boldsymbol{y}^0 & \forall i\in\mathcal{I}\backslash\mathcal{I}^o \\
& \boldsymbol{Y}^\top \boldsymbol{b}_i - \boldsymbol{\beta}_i = \bar{\boldsymbol{\lambda}}_i - \underline{\boldsymbol{\lambda}}_i & \forall i\in\mathcal{I}\backslash\mathcal{I}^o \\
& a_i^0(\boldsymbol{x}) \geq \alpha_i & \forall i\in\mathcal{I} \\
& a_i^j(\boldsymbol{x}) \geq \beta_i^j & \forall i\in\mathcal{I}, j\in\mathcal{J}^+ \\
& a_i^j(\boldsymbol{x}) \leq \beta_i^j & \forall i\in\mathcal{I}, j\in\mathcal{J}^- \\
& a_i^j(\boldsymbol{x}) = \beta_i^j & \forall i\in\mathcal{I}, j\in\mathcal{J} \\
& \boldsymbol{\alpha}\in\mathbb{R}^{|\mathcal{I}|}, \boldsymbol{\beta}\in\mathbb{R}^{|\mathcal{I}|\times I_z}, \boldsymbol{\kappa}\in\mathbb{R}_+^{m+1}, \boldsymbol{r}\in\mathbb{R}^{m+1} & \\
& \boldsymbol{y}^0\in\mathbb{R}^{I_y}, \boldsymbol{Y}\in\mathbb{R}^{I_y\times I_z}, \bar{\boldsymbol{\lambda}},\underline{\boldsymbol{\lambda}}\in\mathbb{R}_+^{|\mathcal{I}\backslash\mathcal{I}^o|\times I_z} &
\end{array}
\qquad (24)
$$

where for any index set $\mathcal{M}$*, we denote* $\bar{\boldsymbol{\alpha}}_{\mathcal{M}} \triangleq \begin{bmatrix} 0 \\ \boldsymbol{\alpha}_{\mathcal{M}} \end{bmatrix} \in \mathbb{R}^{|\mathcal{M}|+1}$, $\bar{\boldsymbol{\beta}}_{\mathcal{M}} \triangleq \begin{bmatrix} \boldsymbol{0}^\top \\ \boldsymbol{\beta}_{\mathcal{M}} \end{bmatrix} \in \mathbb{R}^{(|\mathcal{M}|+1)\times I_z}$, $\bar{\boldsymbol{b}}_{\mathcal{M}} \triangleq \begin{bmatrix} \boldsymbol{0} \ \boldsymbol{b}_{\mathcal{M}} \end{bmatrix} \in \mathbb{R}^{I_y\times(|\mathcal{M}|+1)}$ *and* $\boldsymbol{\alpha}_{\mathcal{M}}$, $\boldsymbol{\beta}_{\mathcal{M}}$, $\boldsymbol{b}_{\mathcal{M}}$ *are the stacked vectors or matrices of* $\alpha_i/\|\boldsymbol{b}_i\|$, $\boldsymbol{\beta}_i^\top/\|\boldsymbol{b}_i\|$, $\boldsymbol{b}_i/\|\boldsymbol{b}_i\|$ *for* $i\in\mathcal{M}$*, respectively.*

4. Multi-period consumption model

We now extend to a T-period problem where uncertainty is revealed at every period and decisions are made dynamically to maximize the total utilities of consumption across all periods. Let $\boldsymbol{x}\in\mathcal{X}$ represent the here-and-now decision. At each period $t\in[T]$, up to I_{ξ_t} independently distributed random factors, *i.e.*, $\tilde{z}_1,...,\tilde{z}_{I_{\xi_t}}$, are realized, where I_{ξ_t} increases with t and $I_{\xi_T} = I_z$. We define the random vector $\tilde{\boldsymbol{\xi}}_t \triangleq (\tilde{z}_1,...,\tilde{z}_{I_{\xi_t}})$ and the vector $\boldsymbol{\xi}_t \triangleq (z_1,...,z_{I_{\xi_t}})$ to represent a realization of $\tilde{\boldsymbol{\xi}}_t$. It is important to note that the set of uncertain factors $\tilde{\boldsymbol{z}}$ is equivalent to $\tilde{\boldsymbol{\xi}}_T$. After $\boldsymbol{\xi}_t$ is realized, the decision-maker makes the t-th period recourse decision $\boldsymbol{y}_t$. We also write $\boldsymbol{y}_t(\boldsymbol{\xi}_t)$ to emphasize the non-anticipative nature of multi-period decision-making, *i.e.*, $\boldsymbol{y}_t$ does not depend on future uncertain outcomes $\tilde{z}_{I_{\xi_t}+1},\ldots,\tilde{z}_{I_z}$.[2] The recourse decision must satisfy the constraint $\boldsymbol{b}_{t,i,\tau}^\top \boldsymbol{y}_\tau(\boldsymbol{\xi}_\tau) \le a_{t,i}^0(\boldsymbol{x}) + \boldsymbol{a}_{t,i}^\top(\boldsymbol{x})\boldsymbol{\xi}_t$ for each $i \in \mathcal{I}_t$ where $\mathcal{I}_t$ represents the set of constraint indices in period t. For each $t \in [T]$, $i \in \mathcal{I}_t$, we assume $\boldsymbol{b}_{t,i,t} \neq \boldsymbol{0}$, and the function $a_{t,i}^j(\boldsymbol{x}) : \mathbb{R}^{I_x} \to \mathbb{R}$ is concave for all $j \in \{0\} \cup \mathcal{J}_t^+$, convex for all $j \in \mathcal{J}_t^-$, and affine for all $j \in \mathcal{J}_t$ where we denote $\mathcal{J}_t^+ \triangleq \mathcal{J}^+ \cap [I_{\xi_t}]$, $\mathcal{J}_t^- \triangleq \mathcal{J}^- \cap [I_{\xi_t}]$, $\mathcal{J}_t \triangleq \mathcal{J} \cap [I_{\xi_t}]$. After making the recourse decision, the decision-maker experiences consumption of $v_t(\boldsymbol{\xi}_t) \triangleq \boldsymbol{c}_t^\top \boldsymbol{y}_t(\boldsymbol{\xi}_t)$. The objective is to maximize the decision-maker's utility of consumption over the entire horizon.

In evaluating the utility of the consumption profile, we adopt the *time-additive* exponential utility preference (Varian 1992, Chapter 19) as follows,

$$u\left(v_1(\boldsymbol{\xi}_1),\ldots,v_T(\boldsymbol{\xi}_T)\right) = \sum_{t\in[T]} \theta_t \left(1 - \exp\left(-\frac{v_t(\boldsymbol{\xi}_t)}{\kappa}\right)\right),$$

where we can specify the temporal discounting via the weights θ_t, $\theta_t \ge 0$. Since we are maximizing the utility, without any loss of generality, we will normalize the weights so that $\sum_{t\in[T]} \theta_t = 1$.

We next generalize the notion of CARA certainty equivalent to the multi-period setting, which we associate with a constant consumption of $v = \mathbb{C}_{\mathbb{P}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]$ at every period so that

$$u(v,\ldots,v) = \mathbb{E}_{\mathbb{P}}\left[u\left(v_1(\tilde{\boldsymbol{\xi}}_1),\ldots,v_T(\tilde{\boldsymbol{\xi}}_T)\right)\right].$$

Definition 4. For a given random vector $\tilde{\boldsymbol{z}} \sim \mathbb{P}$, let $\tilde{\boldsymbol{v}} \triangleq \left(v_1(\tilde{\boldsymbol{\xi}}_1),\ldots,v_T(\tilde{\boldsymbol{\xi}}_T)\right)$ denote the random non-anticipative consumption profile over time. We define the following *multi-period CARA certainty equivalent*

$$\mathbb{C}_{\mathbb{P}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}] \triangleq \begin{cases} \min\limits_{t\in[T]:\theta_t>0} \{\operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}_t]\} & \text{if } \kappa = 0 \\ \sum\limits_{t\in[T]} \theta_t \mathbb{E}_{\mathbb{P}}[\tilde{v}_t] & \text{if } \kappa = \infty \\ -\kappa \log\left(\sum\limits_{t\in[T]} \theta_t \mathbb{E}_{\mathbb{P}}\left[\exp\left(-\dfrac{\tilde{v}_t}{\kappa}\right)\right]\right) & \text{if } \kappa \in (0,\infty). \end{cases}$$

Note that $\mathbb{C}_{\mathbb{P}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]$ coincides with the CARA certainty equivalent $\mathbb{C}_{\mathbb{P}}^{\kappa}[\tilde{v}]$ when $T = 1$.

Proposition 4. *The multi-period CARA certainty equivalent has the following properties:*

1. $\mathbb{C}_{\mathbb{P}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]$ *is non-decreasing in* $\kappa \in [0,\infty]$.

[2] By the standard terminology in multi-period stochastic programming: Let $(\Omega,\mathcal{H},\mathbb{P})$ be a probability space and $\tilde{z}_j : \Omega \to \mathbb{R}$, $j \in [I_z]$ be independent random variables. We define a filtration $\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \cdots \subseteq \mathcal{H}_T = \mathcal{H}$ with $\mathcal{H}_0 = \{\emptyset,\Omega\}$ and $\mathcal{H}_t$ is the σ-algebra generated by $\tilde{z}_1,\cdots,\tilde{z}_{I_{\xi_t}}$. For each $t \in [T]$, we require $\boldsymbol{y}_t$ is adapted to $\mathcal{H}_t$.

2.

$$\min_{t\in[T]:\theta_t>0} \{\operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}_t]\} \leq \mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}] \leq \sum_{t\in[T]} \theta_t \mathbb{E}_{\mathbb{P}}[\tilde{v}_t],$$

and the bounds are achievable.

3. $\mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}]$ *is jointly concave in* $\tilde{\boldsymbol{v}}$ *and* $\kappa > 0$.

4.

$$\begin{aligned} \mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}] = \max_{\boldsymbol{\nu}\in\mathbb{R}^T} & -\kappa \log\left(\sum_{t\in[T]} \theta_t \exp\left(-\frac{\nu_t}{\kappa}\right)\right) \\ \text{s.t.} \quad & \mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_t] \geq \nu_t \qquad \forall t \in [T]. \end{aligned} \tag{25}$$

5. *For all* $\nu \in \mathbb{R}$,

$$\mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}} + \nu\mathbf{1}] = \mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}] + \nu.$$

REMARK 4. The first two properties show the preservation of monotonicity with regards to the risk tolerance level and justify the definition of the multi-period CARA certainty equivalent at its limits. The joint concavity property is also preserved, together with the fourth property showing the connection to CARA certainty equivalent in each period, are essential for tractability of multi-period CARA optimization problems. The last property is the extension of the translation invariance to multi-period so that if each period is increased by the same certain amount, then the multi-period certainty equivalent should also increase by the same amount. This property is sensible and unique to the choice of exponential utility.

In considering ambiguity aversion, it may seem natural to evaluate the worst-case expected cumulative utility for the entire horizon as follows

$$\inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\sum_{t\in[T]} \theta_t \left(1 - \exp\left(-\frac{\tilde{v}_t(\tilde{\boldsymbol{\xi}}_t)}{\kappa}\right)\right)\right].$$

However, we do not know how to tractably evaluate this criterion even if the consumption functions are affinely dependent on the random factors. Instead, we propose to evaluate the Gilboa and Schmeidler (1989) worst-case expected utility for every period in the following criterion,

$$\sum_{t\in[T]} \inf_{\mathbb{P}\in\mathcal{F}} \mathbb{E}_{\mathbb{P}}\left[\theta_t \left(1 - \exp\left(-\frac{\tilde{v}_t(\tilde{\boldsymbol{\xi}}_t)}{\kappa}\right)\right)\right].$$

Apart from the computational benefits, we can also justify this approach as being more prudent in mitigating the ambiguous risk of under consumption or starvation that may occur in any period. Consequently, we propose the following multi-period ambiguity-averse CARA certainty equivalent.

Definition 5. Given an ambiguity set of probability distributions $\mathcal{F}$, we define the *multi-period ambiguity-averse CARA certainty equivalent* as follows

$$\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}] \triangleq \begin{cases} \min_{t\in[T]:\theta_t>0}\left\{\inf_{\mathbb{P}\in\mathcal{F}} \operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}_t]\right\} & \text{if } \kappa=0 \\ \sum_{t\in[T]}\theta_t \inf_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}[\tilde{v}_t] & \text{if } \kappa=\infty \\ -\kappa\log\left(\sum_{t\in[T]}\theta_t\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\exp\left(-\frac{\tilde{v}_t}{\kappa}\right)\right]\right) & \text{if } \kappa\in(0,\infty)\end{cases}$$

where $\tilde{\boldsymbol{v}} \triangleq \left(\boldsymbol{c}_1^\top\boldsymbol{y}_1(\tilde{\boldsymbol{\xi}}_1),\ldots,\boldsymbol{c}_T^\top\boldsymbol{y}_T(\tilde{\boldsymbol{\xi}}_T)\right)$ and $\tilde{\boldsymbol{\xi}}_T\sim\mathbb{P}\in\mathcal{F}$.
Note that $\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]$ is well-defined under Assumption 1 if $\boldsymbol{y}_t\in\mathcal{R}^{I_{\xi_t},I_{y_t}}$ for each $t\in[T]$.

Proposition 5. *The multi-period ambiguity-averse CARA certainty equivalent has the following properties:*

1. $\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]$ *is non-decreasing in* $\kappa\in[0,\infty]$.
2. $$\min_{t\in[T]:\theta_t>0}\left\{\inf_{\mathbb{P}\in\mathcal{F}}\operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}_t]\right\}\le\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]\le\sum_{t\in[T]}\theta_t\inf_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}[\tilde{v}_t],$$
   and the limits are achievable.
3. $\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]$ *is jointly concave in* $\tilde{\boldsymbol{v}}$ *and* $\kappa>0$.
4. $$\begin{aligned}\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]=\max_{\boldsymbol{\nu}\in\mathbb{R}^T}\ & -\kappa\log\left(\sum_{t\in[T]}\theta_t\exp(-\nu_t/\kappa)\right) \\ \text{s.t.}\ & \mathbb{C}_{\mathcal{F}}^{\kappa}[\tilde{v}_t]\ge\nu_t \qquad \forall t\in[T].\end{aligned} \tag{26}$$
5. *For all* $\nu\in\mathbb{R}$,
   $$\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}+\nu\mathbf{1}]=\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]+\nu.$$
6. $$\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}]\le\inf_{\mathbb{P}\in\mathcal{F}}\mathbb{C}_{\mathbb{P}}^{\kappa,\boldsymbol{\theta}}[\tilde{\boldsymbol{v}}].$$

Remark 5. We see the multi-period ambiguity-averse CARA certainty equivalent preserves the salient properties of multi-period CARA certainty equivalent in Proposition 4. The last property shows that the multi-period ambiguity-averse CARA certainty equivalent is a more conservative (or robust) evaluation of the worst-case achievable multi-period CARA certainty equivalent evaluated at the beginning of the time horizon.

We are now ready to propose our robust CARA multi-period consumption model as follows,

$$\begin{aligned}\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T}\ & \mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}\left[\boldsymbol{c}_1^\top\boldsymbol{y}_1(\tilde{\boldsymbol{\xi}}_1),\ldots,\boldsymbol{c}_T^\top\boldsymbol{y}_T(\tilde{\boldsymbol{\xi}}_T)\right] \\ \text{s.t.}\ & \sum_{\tau\in[t]}\boldsymbol{b}_{t,i,\tau}^\top\boldsymbol{y}_\tau(\boldsymbol{\xi}_\tau)\le a_{t,i}^0(\boldsymbol{x})+\boldsymbol{a}_{t,i}^\top(\boldsymbol{x})\boldsymbol{\xi}_t \qquad \forall t\in[T],\ \forall i\in\mathcal{I}_t,\ \forall\boldsymbol{z}\in\mathcal{Z} \\ & \boldsymbol{y}_t\in\mathcal{R}^{I_{\xi_t},I_{y_t}} \qquad \forall t\in[T].\end{aligned} \tag{27}$$

Observe that Problem (27) generalizes the robust CARA optimization with two-stage payoff function (16). We can also consider maximizing the expected CARA utility of the total payoffs in the following multi-period optimization model,

$$
\begin{aligned}
\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{y}_1,\dots,\boldsymbol{y}_T} \quad & \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\sum_{t\in[T]} \boldsymbol{c}_t^\top \boldsymbol{y}_t(\tilde{\boldsymbol{\xi}}_t)\right] \\
\text{s.t.} \quad & \sum_{\tau\in[t]} \boldsymbol{b}_{t,i,\tau}^\top \boldsymbol{y}_\tau(\boldsymbol{\xi}_\tau) \le a_{t,i}^0(\boldsymbol{x}) + \boldsymbol{a}_{t,i}^\top(\boldsymbol{x})\boldsymbol{\xi}_t \quad \forall t\in[T],\ \forall i\in\mathcal{I}_t,\ \forall \boldsymbol{z}\in\mathcal{Z} \\
& \boldsymbol{y}_t \in \mathcal{R}^{I_{\xi_t}, I_{y_t}} \quad \forall t\in[T].
\end{aligned} \tag{28}
$$

Incidentally, this is also a special case of the T-period consumption model of Problem (27) with $\theta_t = 0$ for all $t\in[T-1]$, $\theta_T = 1$, an auxiliary recourse decision $y_{T,I_{y_T}+1}$ in period T as the consumption, and one more constraint $y_{T,I_{y_T}+1}(\boldsymbol{\xi}_T) - \sum_{t\in[T]} \boldsymbol{c}_t^\top \boldsymbol{y}_t(\boldsymbol{\xi}_t) \le 0$ for all $\boldsymbol{z}\in\mathcal{Z}$.

We next extend the MLDR to Problem (27) with non-anticipativity consideration. For this purpose, we first solve for each $t\in[T]$, $i\in\mathcal{I}_t$, the problem

$$
\begin{aligned}
\max_{\boldsymbol{y}_t,\dots,\boldsymbol{y}_T} \quad & \sum_{\tau=t}^{T} \theta_\tau \left(1-\exp\left(-\boldsymbol{c}_\tau^\top \boldsymbol{y}_\tau/\kappa\right)\right) \\
\text{s.t.} \quad & \sum_{\tau=t}^{s} \boldsymbol{b}_{s,k,\tau}^\top \boldsymbol{y}_\tau \le 0 \quad && \forall s\in\{t,\dots,T\}, k\in\mathcal{I}_s \\
& \boldsymbol{b}_{t,i,t}^\top \boldsymbol{y}_t = -\|\boldsymbol{b}_{t,i,t}\| \\
& \boldsymbol{c}_\tau^\top \boldsymbol{y}_\tau \le 0 && \forall \tau\in\{t,\dots,T\} \\
& \boldsymbol{y}_\tau \in \mathbb{R}^{I_\tau} && \forall \tau\in\{t,\dots,T\}.
\end{aligned} \tag{29}
$$

For each $t\in[T]$, we denote $\mathcal{I}_t^o \subseteq \mathcal{I}_t$ as the index set of i such that the above optimization problem is feasible and $\boldsymbol{y}_{\tau*}^{t,i}$ as the corresponding optimal solution for any $\tau\in\{t,\dots,T\}$. Similar to the two-stage case, we can further partition the index set $\mathcal{I}_t^o$ as $\mathcal{I}_t^o = \bigcup_{\ell\in[m_t]} \mathcal{I}_{t,\ell}^o$ such that $\boldsymbol{y}_{\tau*}^{t,i_1} = \boldsymbol{y}_{\tau*}^{t,i_2}$ if and only if i_1 and i_2 are in the same $\mathcal{I}_{t,\ell}^o$ and denote $\boldsymbol{y}_{\tau*}^{t,\ell}$ as any $\boldsymbol{y}_{\tau*}^{t,i}$ with $i\in\mathcal{I}_{t,\ell}^o$. Subsequently, we propose the multi-period MLDR

$$
\hat{\boldsymbol{y}}_t(\boldsymbol{\xi}_t) \triangleq \bar{\boldsymbol{y}}_t(\boldsymbol{\xi}_t) + \sum_{s\in[t]}\sum_{\ell\in[m_s]} \boldsymbol{y}_{t*}^{s,\ell}\left(\max_{i\in\mathcal{I}_{s,\ell}^o}\left\{\frac{h_{s,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[s]}(\boldsymbol{\xi}_s),\boldsymbol{\xi}_s)}{\|\boldsymbol{b}_{s,i,s}\|}\right\}\right)^+ \tag{30}
$$

where for each $t\in[T]$,

$$
\begin{aligned}
\bar{\boldsymbol{y}}_t(\boldsymbol{\xi}_t) &\triangleq \boldsymbol{y}_t^0 + \boldsymbol{Y}_t\boldsymbol{\xi}_t \\
h_{t,i}(\boldsymbol{x},\boldsymbol{y}_{[t]},\boldsymbol{\xi}_t) &\triangleq \sum_{\tau\in[t]} \boldsymbol{b}_{t,i,\tau}^\top \boldsymbol{y}_\tau - a_{t,i}^0(\boldsymbol{x}) - \boldsymbol{a}_{t,i}^\top(\boldsymbol{x})\boldsymbol{\xi}_t
\end{aligned}
$$

with $\boldsymbol{y}_t^0 \in \mathbb{R}^{I_{y_t}}$, $\boldsymbol{Y}_t \triangleq [\boldsymbol{y}_t^1,\dots,\boldsymbol{y}_t^{I_{\xi_t}}] \in \mathbb{R}^{I_{y_t}\times I_{\xi_t}}$, and $\boldsymbol{y}_{[t]}$ is the collection of $\boldsymbol{y}_\tau$ for $\tau\in[t]$.

We establish the feasibility of the multi-period MLDR as follows.

Proposition 6. *For each $t \in [T]$, the multi-period MLDR, $\hat{\boldsymbol{y}}_t$ satisfies the non-anticipativity constraints. Moreover, suppose $\bar{\boldsymbol{y}}_t \in \mathcal{L}^{I_{\xi_t}, I_{y_t}}$ satisfies $\sum_{\tau \in [t]} \boldsymbol{b}_{t,i,\tau}^\top \bar{\boldsymbol{y}}_\tau(\boldsymbol{\xi}_\tau) \leq a_{t,i}^0(\boldsymbol{x}) + \boldsymbol{a}_{t,i}^\top(\boldsymbol{x})\boldsymbol{\xi}_t$ for each $i \in \mathcal{I}_t \backslash \mathcal{I}_t^o$, then $\hat{\boldsymbol{y}}_t$ is feasible to Problem (27).*

We construct a tractable approximation of Problem (27) as follows.

Theorem 8. *The multi-period model (27) has a practicably solvable lower bound as follows:*

$$
\begin{aligned}
\max_{\substack{\rho, \boldsymbol{x}, p_t, \nu_t, \boldsymbol{\alpha}_t, \\ \boldsymbol{\beta}_t, \boldsymbol{y}_t^0, \boldsymbol{Y}_t, \boldsymbol{\kappa}^t, \boldsymbol{r}^t, \\ \bar{\boldsymbol{\lambda}}_t, \underline{\boldsymbol{\lambda}}_t, \forall t \in [T]}} \quad & \rho \\
\text{s.t.} \quad & \sum_{t \in [T]} \theta_t p_t \leq \kappa \\
& (p_t, \kappa, \rho - \nu_t) \in \mathcal{K}_{\exp} && \forall t \in [T] \\
& r_0^t - \sum_{s \in [t]} \sum_{\ell \in [m_s]} (\boldsymbol{c}_t^\top \boldsymbol{y}_{t*}^{s,\ell}) r_{s,\ell}^t \geq \nu_t && \forall t \in [T] \\
& \kappa_0^t - \sum_{s \in [t]} \sum_{\ell \in [m_s]} (\boldsymbol{c}_t^\top \boldsymbol{y}_{t*}^{s,\ell}) \kappa_{s,\ell}^t = \kappa && \forall t \in [T] \\
& \boldsymbol{c}_t^\top \boldsymbol{y}_t^0 + \sum_{j \in [I_{\xi_t}]} \phi_j(\kappa_0^t, \boldsymbol{c}_t^\top \boldsymbol{y}_t^j) \geq r_0^t && \forall t \in [T] \\
& \Phi\left(\kappa_{s,\ell}^t,\ \bar{\boldsymbol{\alpha}}_{s,\mathcal{I}_{s,\ell}^o} - \sum_{\tau \in [s]} \bar{\boldsymbol{b}}_{s,\mathcal{I}_{s,\ell}^o,\tau}^\top \boldsymbol{y}_\tau^0,\ \bar{\boldsymbol{\beta}}_{s,\mathcal{I}_{s,\ell}^o} - \sum_{\tau \in [s]} \bar{\boldsymbol{b}}_{s,\mathcal{I}_{s,\ell}^o,\tau}^\top \bar{\boldsymbol{Y}}_{\tau s}\right) \geq r_{s,\ell}^t && \forall t \in [T], s \in [t], \ell \in [m_s] \\
& \bar{\boldsymbol{\lambda}}_{t,i}^\top \bar{\boldsymbol{\xi}}_t - \underline{\boldsymbol{\lambda}}_{t,i}^\top \underline{\boldsymbol{\xi}}_t \leq \alpha_{t,i} - \sum_{\tau \in [t]} \boldsymbol{b}_{t,i,\tau}^\top \boldsymbol{y}_\tau^0 && \forall t \in [T], i \in \mathcal{I}_t \backslash \mathcal{I}_t^o \\
& \sum_{\tau \in [t]} \bar{\boldsymbol{Y}}_{\tau t}^\top \boldsymbol{b}_{t,i,\tau} - \boldsymbol{\beta}_{t,i} = \bar{\boldsymbol{\lambda}}_{t,i} - \underline{\boldsymbol{\lambda}}_{t,i} && \forall t \in [T], i \in \mathcal{I}_t \backslash \mathcal{I}_t^o \\
& a_{t,i}^0(\boldsymbol{x}) \geq \alpha_{t,i} && \forall t \in [T], i \in \mathcal{I}_t \\
& a_{t,i}^j(\boldsymbol{x}) \geq \beta_{t,i}^j && \forall t \in [T], i \in \mathcal{I}_t, j \in \mathcal{J}_t^+ \\
& a_{t,i}^j(\boldsymbol{x}) \leq \beta_{t,i}^j && \forall t \in [T], i \in \mathcal{I}_t, j \in \mathcal{J}_t^- \\
& a_{t,i}^j(\boldsymbol{x}) = \beta_{t,i}^j && \forall t \in [T], i \in \mathcal{I}_t, j \in \mathcal{J}_t \\
& \rho \in \mathbb{R}, \boldsymbol{x} \in \mathcal{X} \\
& p_t, \nu_t \in \mathbb{R}, \boldsymbol{\alpha}_t \in \mathbb{R}^{|\mathcal{I}_t|}, \boldsymbol{\beta}_t \in \mathbb{R}^{|\mathcal{I}_t| \times I_{\xi_t}}, \boldsymbol{y}_t^0 \in \mathbb{R}^{I_{y_t}}, \boldsymbol{Y}_t \in \mathbb{R}^{I_{y_t} \times I_{\xi_t}} && \forall t \in [T] \\
& \boldsymbol{\kappa}^t \in \mathbb{R}_+^{1+\sum_{s \in [t]} m_s}, \boldsymbol{r}^t \in \mathbb{R}^{1+\sum_{s \in [t]} m_s}, \bar{\boldsymbol{\lambda}}_t, \underline{\boldsymbol{\lambda}}_t \in \mathbb{R}_+^{|\mathcal{I}_t \backslash \mathcal{I}_t^o| \times I_{\xi_t}} && \forall t \in [T]
\end{aligned}
\tag{31}
$$

where $\bar{\boldsymbol{Y}}_{\tau t} \triangleq \begin{bmatrix} \boldsymbol{Y}_\tau & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{I_{y_\tau} \times I_{\xi_t}}$ *for* $\tau \leq t$, $\bar{\boldsymbol{\alpha}}_{s,\mathcal{I}_{s,\ell}^o} \triangleq \begin{bmatrix} 0 \\ \boldsymbol{\alpha}_{s,\mathcal{I}_{s,\ell}^o} \end{bmatrix} \in \mathbb{R}^{|\mathcal{I}_{s,\ell}^o|+1}$, $\bar{\boldsymbol{\beta}}_{s,\mathcal{I}_{s,\ell}^o} \triangleq \begin{bmatrix} \boldsymbol{0}^\top \\ \boldsymbol{\beta}_{s,\mathcal{I}_{s,\ell}^o} \end{bmatrix} \in \mathbb{R}^{(|\mathcal{I}_{s,\ell}^o|+1) \times I_{\xi_s}}$, $\bar{\boldsymbol{b}}_{s,\mathcal{I}_{s,\ell}^o,\tau} \triangleq \begin{bmatrix} \boldsymbol{0} & \boldsymbol{b}_{s,\mathcal{I}_{s,\ell}^o,\tau} \end{bmatrix} \in \mathbb{R}^{I_{y_\tau} \times (|\mathcal{I}_{s,\ell}^o|+1)}$ *and* $\boldsymbol{\alpha}_{s,\mathcal{I}_{s,\ell}^o}$, $\boldsymbol{\beta}_{s,\mathcal{I}_{s,\ell}^o}$, $\boldsymbol{b}_{s,\mathcal{I}_{s,\ell}^o,\tau}$ *are the stacked vectors or matrices of* $\alpha_{s,i}/\|\boldsymbol{b}_{s,i,s}\|$, $\boldsymbol{\beta}_{s,i}^\top/\|\boldsymbol{b}_{s,i,s}\|$, $\boldsymbol{b}_{s,i,\tau}/\|\boldsymbol{b}_{s,i,s}\|$ *for* $s \in [T]$, $\ell \in [m_s]$, $i \in \mathcal{I}_{s,\ell}^o$, *respectively.*

Note that in practical implementation of the robust optimization solutions, we may ignore the solutions for the MLDR, but only to implement the solutions for the here-and-now decision,

$\boldsymbol{x} \in \mathcal{X}$. As uncertainty unfolds, the future wait-and-see decision will become here-and-now. In a rolling-horizon implementation, this decision can be obtained by solving a new robust optimization problem with updated priors (see, *e.g.,* Ben-Tal et al. 2004, Bertsimas et al. 2019).

On time consistency

We are aware that time consistency is a *maxim* in many multi-period stochastic programming models mandating that an optimal policy perceived in one time period must be recognized as optimal in another. However, we do not enforce time consistency as a preference in our multi-period decision framework. Apart from the time inconsistency issues that may arise from robust decision making (see, *e.g.,* Delage and Iancu 2015), we also allow arbitrary choice of temporal discounting, such as behaviorally inspired *hyperbolic discounting* (Laibson 1997) that would result in time inconsistent preferences. In reality, time consistency is not a dominant human behavior even in the absence of uncertainty (see, *e.g.*, Loch and Wu 2007, Frederick et al. 2002). Moreover, since dynamic optimization problems are potentially PSPACE-hard (Dyer and Stougie 2006), the consideration of time consistency presupposes an impractical amount of computational resources needed to ensure the optimality of a time consistent policy. We acknowledge that our model may not cater to a fully rational agent with unlimited computational resources.

Our stand to relegate time consistency is not uncommon in the literature, and we refer interested readers to Kydland and Prescott (1977), Bajeux-Besnainou and Portait (1998). A trivial fix is to adopt the *pre-committed policy* approach by firmly adhering to the optimum policy evaluated at the first period throughout the planning horizon. In one of our numerical studies, we evaluate this approach by comparing the performance of the pre-committed MLDR policy against the time-consistent optimal DP policy, with both policies constructed from an empirical distribution at the beginning of the period. Another common criticism of robust optimization is the perceived over-conservativeness, which may not be true with more sophisticated ambiguity sets and approximation techniques such as those introduced in Goh and Sim (2010), See and Sim (2010), Chen et al. (2020). The proof of the pudding should be in its eating. Hence, it is imperative for us to compare the quality of the solutions obtained from the deterministic approximations of our robust CARA optimization models against those obtained from the Monte-Carlo approximations of stochastic CARA optimization models.

5. Numerical studies

In this section, we apply the tractable approximation of robust CARA optimization models to study its numerical performance on solving two adaptive linear optimization problems. In the first experiment, we consider a project management problem, and benchmark our solutions against those obtained from SAA approximations of stochastic optimization. In the second experiment, we study a multi-period inventory management problem and benchmark the multi-period MLDR policy against the policy obtained using dynamic programming (DP). In both problems we show that our tractable approximation yields solutions with better out-of-sample performance when there are insufficient training samples or when the risk tolerance levels are low.

Project management

We consider solving a risk-averse project management problem (*e.g.* Ben-Tal et al. 2009, Chen et al. 2007b) via our tractable approximation and a stochastic optimization model using an empirical distribution to mimic a data-driven setting where the underlying data generating model is not known to the decision maker. We fix the empirical distribution and vary the risk tolerance level to obtain the solution profiles of both approaches.

To represent the project management problem, we use a directed acyclic graph with n nodes and m arcs, denoted by $\mathcal{E}$. Each node corresponds to an event that signals the completion of a subset of activities, while each arc corresponds to an activity linking two events. An event only occurs when all the activities that correspond to its incoming edges have been completed. We use node 1 as the start event and the last node n as the end event. The completion time of event $i \in [n]$ is denoted by y_i. Each activity $(i,j) \in \mathcal{E}$ is associated with an uncertain processing time $\tilde{t}_{ij}$, which starts only after the event corresponding to the originating node has occurred. We assume that the random processing time $\tilde{t}_{ij}$ depends on the allocated additional resources and can be represented as $\tilde{t}_{ij} = b_{ij} + a_{ij}\tilde{z}_{ij}(1 - x_{ij})$, where $\tilde{z}_{ij}$ is a zero-mean random variable, and $x_{ij} \in [0, \bar{x}_{ij}]$ is the amount of resources allocated to activity $(i,j) \in \mathcal{E}$. We assume that the random processing times $\tilde{t}_{ij}$ are independent of each other, and are non-negative for all realizations of $\tilde{z}_{ij}$ and all ranges of x_{ij}. We denote c_{ij} as the cost of using each unit of resource for the activity on the arc (i,j).

Our goal is to determine an optimal resource allocation decision $\boldsymbol{x}$ that minimizes the CARA certainty equivalent of the completion time of the project, subject to the constraint that the total available resources do not exceed a budget C, *i.e.*,

$$\min_{\boldsymbol{x} \in \mathcal{X}} \overline{\mathbb{C}}^{\kappa}_{\mathbb{P}} \left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}}) \right]$$

where

$$\begin{aligned} f(\boldsymbol{x}, \boldsymbol{z}) = \min_{\boldsymbol{y}} \;& y_n \\ \text{s.t.} \;& y_j \geq y_i + b_{ij} + a_{ij} z_{ij}(1 - x_{ij}) \;\; \forall (i,j) \in \mathcal{E}\,, \\ & y_1 = 0 \end{aligned} \tag{32}$$

and

$$\mathcal{X} = \left\{ \boldsymbol{x} : \; \boldsymbol{c}^\top \boldsymbol{x} \leq C, \; \boldsymbol{0} \leq \boldsymbol{x} \leq \bar{\boldsymbol{x}} \right\}.$$

Our computational setup follows Chen et al. (2008), where we consider a fictitious activity network represented by an H-by-W grid, with a total of $n = H \times W$ nodes and $m = H(W-1) + W(H-1)$ arcs. The first node is located at the bottom left corner, while the last node is at the upper right corner. Each arc on the graph proceeds either towards the right node or the upper node. We set $H = 4$ and $W = 6$, resulting in $n = 24$ and $m = 38$. For all activities, we assume $a_{ij} = b_{ij} = 3$, $\bar{x}_{ij} = 1$, and $c_{ij} = 1$. The processing times are generated from stochastically independent factors, and each factor $\tilde{z}_{ij}$ is a two-point random variable that takes on the value of $1/(2\beta)$ with probability β or $-1/(2(1-\beta))$ with probability $1-\beta$, where $\beta \in \{0.1, 0.2, 0.4\}$ controls the variance of the random duration. However, the distribution is unknown to the decision maker. Instead, we generate $S \in \{20, 10000\}$ i.i.d. samples of $\tilde{\boldsymbol{z}}$ from the underlying distribution for each problem instance and we denote each sample by $\hat{\boldsymbol{z}}^s$, $s \in [S]$. Then we obtain the here-and-now decisions by an SAA approach and an MLDR-based approximation approach, denoted as $\boldsymbol{x}^S$ and $\boldsymbol{x}^M$, respectively. For SAA, we use the empirical average to replace the expectation in the certainty equivalent, *i.e.*,

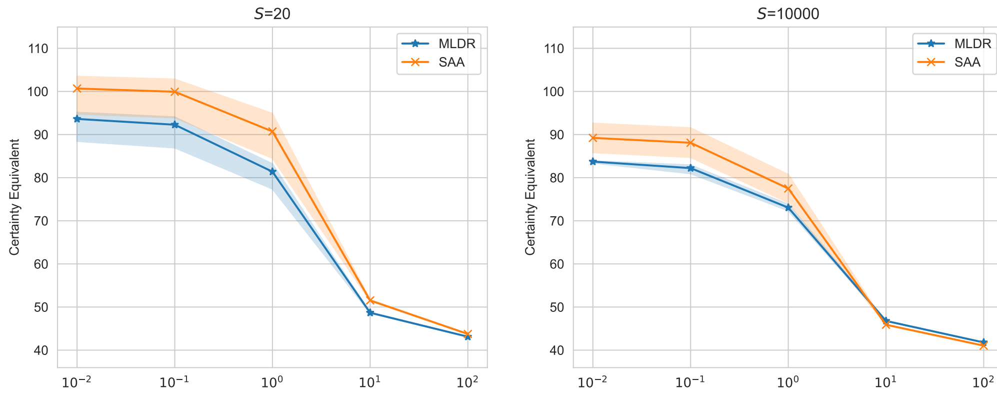
$$\min_{\boldsymbol{x} \in \mathcal{X}} \kappa \log \frac{1}{S} \sum_{s \in [S]} \exp\left(\frac{f(\boldsymbol{x}, \hat{\boldsymbol{z}}^s)}{\kappa}\right).$$

For the MLDR-based approximation, we solve $\overline{\mathbb{C}}^{\kappa}_{\mathcal{F}}[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})]$ where the ambiguity set $\mathcal{F}$ is based on mean, support and mean absolute deviation estimated from data. We emphasize that MLDR is only used for approximating the ambiguity-averse CARA certainty equivalent. After obtaining $\boldsymbol{x}^S$ and $\boldsymbol{x}^M$, we implement them and evaluate the out-of-sample certainty equivalent of $f(\boldsymbol{x}, \tilde{\boldsymbol{z}})$ on 50,000 i.i.d. test samples generated from the underlying distribution. We set risk tolerance parameter $\kappa \in \{0.01, 0.1, 1, 10, 100\}$ and budget $C = 12$. The results are averaged over 100 random problem instances. In Figure 1, we plot the *out-of-sample* CARA certainty equivalents of the project completion times under different risk tolerance parameters. We have also tested with $C \in \{4, 20\}$ and the results are similar.

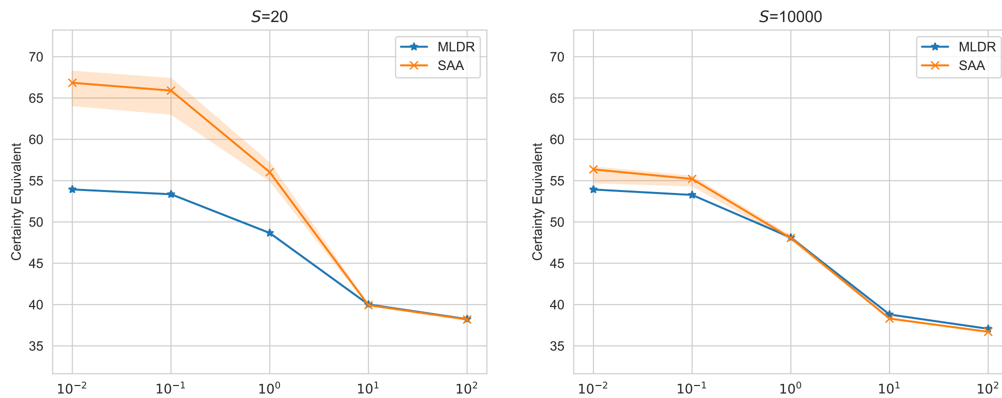
We observe that when the training sample size is limited ($S = 20$), the solution $\boldsymbol{x}^M$ outperforms $\boldsymbol{x}^S$. Furthermore, the gap in out-of-sample certainty equivalent between the two solutions becomes larger as κ decreases. As expected, the out-of-sample certainty equivalent evaluated at SAA solutions improves as S increases from 20 to 10,000. With a very large training size ($S = 10,000$), $\boldsymbol{x}^S$ performs slightly better than $\boldsymbol{x}^M$ when $\kappa \in \{10, 100\}$; however, it is still dominated by $\boldsymbol{x}^M$ when the risk tolerance level is low and the variance is high. Therefore, for a fixed sample size, it becomes more challenging for the SAA approach to maintain the quality of the approximation as the risk tolerance level decreases. Moreover, we find that the out-of-sample certainty equivalent of $\boldsymbol{x}^M$ is more concentrated than that of $\boldsymbol{x}^S$, showing the robustness of $\boldsymbol{x}^M$ over $\boldsymbol{x}^S$, especially when the random duration has a high variance ($\beta = 0.1$) and the risk tolerance level is low. Finally, we note

Figure 1 Certainty equivalent of completion time under different risk tolerance parameters

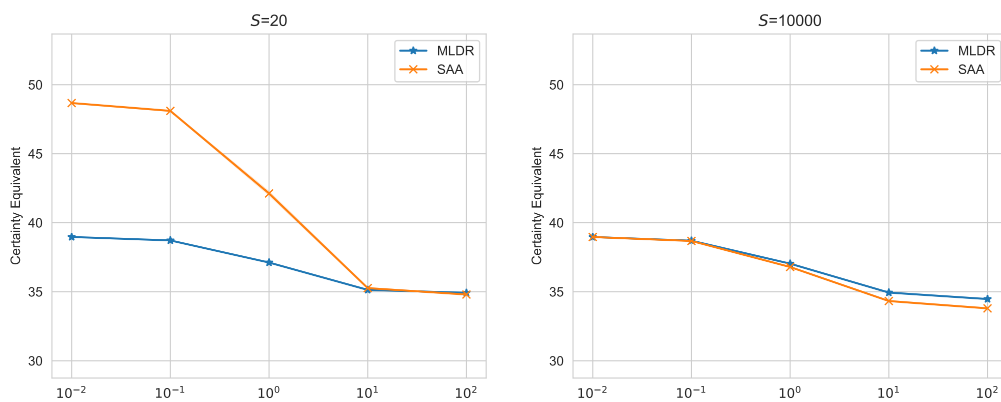
(a) $\beta = 0.1$



(b) $\beta = 0.2$



(c) $\beta = 0.4$



Notes. We visualize the tube between the 10% and 90% quantiles (shaded areas) as well as the mean value (solid lines) of the out-of-sample performance over 100 random instances. Some shaded areas may not be visible in the plot due to the concentration of certain equivalents.

that the average computation time of the MLDR-based approach is 1.79 seconds (independent of the sample size S), while the average computation time of SAA is 0.05 seconds when $S = 20$ and 11.49 seconds when $S = 10,000$, respectively. We see that, unlike the SAA approach, the MLDR-based approach does not suffer from increasing computational burden with larger training sample sizes.

Based on this experiment, we can conclude that our tractable approximation is a viable alternative to the SAA approach. It is particularly advantageous in situations where a low-risk tolerance level is desired and when the training data is limited.

Multi-period inventory management with financing

In the second experiment, we apply the multi-period MLDR approximation approach to solve a risk-averse multi-period inventory management with financing problem. We benchmark its performance with the optimum policy obtained via dynamic programming (DP) using a limited sized empirical distribution that is sampled from the true distribution. In contrast to the previous experiment, we will implement the multi-period MLDR as a pre-committed policy in our numerical study, which is less ideal than a rolling-horizon implementation, but will greatly accelerate our computational studies. A similar robust optimization model has also been proposed in See and Sim (2010) to address a risk-neutral multi-period inventory management without the consideration of financing. Unfortunately, their proposed approximations do not naturally extend to the CARA criterion.

Specifically, we consider a multi-period inventory management problem proposed in Chen et al. (2007a), where the risk-averse firm determines the inventory and financing policies to maximize the expected utility of consumption over a finite time horizon. We assume the demand is exogenous and stochastically independent across periods. At the beginning of each period $t \in [T]$, the inventory level is x_t; the firm makes a replenishment decision $y_t \geq x_t$ at the cost of $c_t(y_t - x_t)$ before the uncertain demand $\tilde{z}_t$ is realized. We assume the unsatisfied demand is backlogged so that the next-period inventory level is $x_{t+1} = y_t - \tilde{z}_t$ and the firm obtains an income $q_t = p_t\tilde{z}_t - h(y_t - \tilde{z}_t)^+ - b(\tilde{z}_t - y_t)^+ - c_t(y_t - x_t)$, where p_t is the unit selling price, h is the unit holding cost, and b is the unit backlogging cost. Subsequently, the firm determines the consumption level f_t and receives the corresponding utility $\left(1 - e^{-f_t/\kappa}\right)$. We do not consider temporal discounting and set $\theta_t = 1/T,\ t \in [T]$. The consumption decision also determines the financing decision so that its wealth w_t transits according to $w_{t+1} = (1+\beta)(w_t + q_t - f_t)$ where β is the interest rate. We can interpret β as either the saving or borrowing rate depending on whether $(w_t + q_t - f_t)$ is positive or negative, respectively. To tractably solve this problem by DP, it is necessary to assume that the saving and borrowing rates are identical (Chen et al. 2007a). We assume the firm aims to maximize the expected time-additive exponential utility function of consumption so the problem can be formulated as

$$
\begin{array}{lll}
\max\limits_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{f},\boldsymbol{w},\boldsymbol{q}} & \mathbb{E}_{\mathbb{P}}\left[\sum\limits_{t\in[T]}\theta_t(1-e^{-f_t(\tilde{\boldsymbol{\xi}}_t)/\kappa})\right] & \\
\text{s.t.} & f_t(\tilde{\boldsymbol{\xi}}_t)=w_t(\tilde{\boldsymbol{\xi}}_{t-1})-\frac{w_{t+1}(\tilde{\boldsymbol{\xi}}_t)}{1+\beta}+q_t(\tilde{\boldsymbol{\xi}}_t) & \forall t\in[T]\\
& q_t(\tilde{\boldsymbol{\xi}}_t)\le p_t\tilde{z}_t-h(y_t(\tilde{\boldsymbol{\xi}}_{t-1})-\tilde{z}_t)-c_t(y_t(\tilde{\boldsymbol{\xi}}_{t-1})-x_t(\tilde{\boldsymbol{\xi}}_{t-1})) & \forall t\in[T]\\
& q_t(\tilde{\boldsymbol{\xi}}_t)\le p_t\tilde{z}_t-b(\tilde{z}_t-y_t(\tilde{\boldsymbol{\xi}}_{t-1}))-c_t(y_t(\tilde{\boldsymbol{\xi}}_{t-1})-x_t(\tilde{\boldsymbol{\xi}}_{t-1})) & \forall t\in[T]\\
& y_t(\tilde{\boldsymbol{\xi}}_{t-1})\ge x_t(\tilde{\boldsymbol{\xi}}_{t-1}) & \forall t\in[T]\\
& x_{t+1}(\tilde{\boldsymbol{\xi}}_t)=y_t(\tilde{\boldsymbol{\xi}}_{t-1})-\tilde{z}_t & \forall t\in[T-1]\\
& w_{T+1}(\tilde{\boldsymbol{\xi}}_T)=0 &
\end{array}
\tag{33}
$$

where the initial wealth w_1 and inventory level x_1 are given. Chen et al. (2007a) shows that a base-stock policy is optimal for all $t\in[T]$. Moreover, the optimal policy of problem (33) can be obtained by solving the DP with Bellman equation

$$
G_t(x)=\max_{y\ge x}\mathbb{C}_{\mathbb{P}}^{R_t}\left[q_t(y,\tilde{z}_t)+\frac{1}{1+\beta}G_{t+1}(y-\tilde{z}_t)\right]
$$

where the *effective risk tolerance* $R_t=\sum_{\tau=t}^{T}\frac{\kappa}{(1+\beta)^{\tau-t}}$ and $G_{T+1}(x)=0$. The optimal consumption is given by

$$
f_t^*(w,y,\tilde{z})=\frac{\kappa}{R_t}\left(w+q_t(y,\tilde{z})+\frac{1}{1+\beta}G_{t+1}(y-\tilde{z})\right)+C_t
$$

where $C_t=-\frac{R_{t+1}\kappa}{R_t(1+\beta)}\log\frac{A_{t+1}(1+\beta)\kappa}{\theta_tR_{t+1}}$ and $A_t=\frac{(1+\beta)R_t}{R_{t+1}}A_{t+1}\left(\frac{A_{t+1}(1+\beta)\kappa}{\theta_tR_{t+1}}\right)^{-\kappa/R_t}$.

We also solve the problem (33) using our multi-period MLDR approximation approach. To do so, we first have to remove the equality constraints of the problem; we eliminate x_t and q_t by substitution. The reformulated problem becomes,

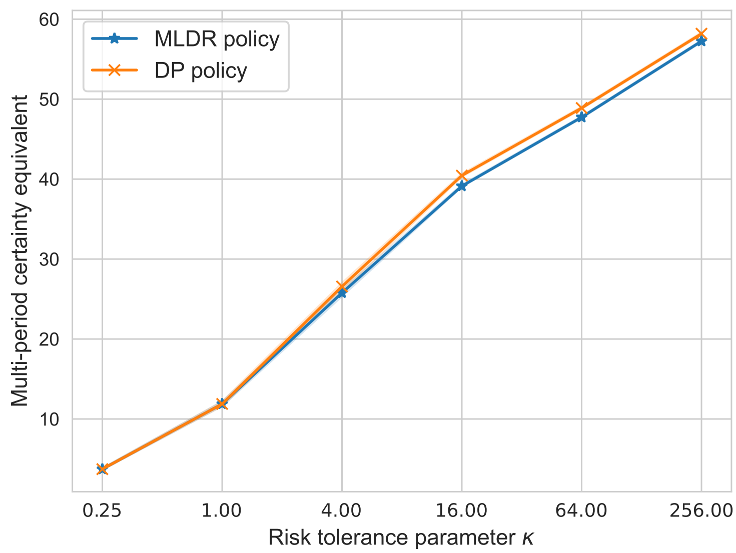
$$
\begin{aligned}
\max_{\boldsymbol{y},\boldsymbol{f},\boldsymbol{w}} \ & \mathbb{C}_{\mathbb{P}}^{\kappa,\boldsymbol{\theta}}\left[\boldsymbol{f}(\tilde{\boldsymbol{z}})\right] \\
\text{s.t. } & y_1 \geq x_1 \\
& y_t(\tilde{\boldsymbol{\xi}}_{t-1}) \geq y_{t-1}(\tilde{\boldsymbol{\xi}}_{t-2}) - \tilde{z}_{t-1} && \forall t \in \{2,...,T\} \\
& f_1(\tilde{\boldsymbol{\xi}}_1) \leq w_1 - \tfrac{w_2(\tilde{\boldsymbol{\xi}}_1)}{1+\beta} + (p_1+h)\tilde{z}_1 - (h+c_1)y_1 + c_1 x_1 \\
& f_1(\tilde{\boldsymbol{\xi}}_1) \leq w_1 - \tfrac{w_2(\tilde{\boldsymbol{\xi}}_1)}{1+\beta} + (p_1-b)\tilde{z}_1 + (b-c_1)y_1 + c_1 x_1 \\
& f_t(\tilde{\boldsymbol{\xi}}_t) \leq w_t(\tilde{\boldsymbol{\xi}}_{t-1}) - \tfrac{w_{t+1}(\tilde{\boldsymbol{\xi}}_t)}{1+\beta} + (p_t+h)\tilde{z}_t - c_t\tilde{z}_{t-1} \\
& \qquad\qquad -(h+c_t)y_t(\tilde{\boldsymbol{\xi}}_{t-1}) + c_t y_{t-1}(\tilde{\boldsymbol{\xi}}_{t-2}) && \forall t \in \{2,...,T-1\} \\
& f_t(\tilde{\boldsymbol{\xi}}_t) \leq w_t(\tilde{\boldsymbol{\xi}}_{t-1}) - \tfrac{w_{t+1}(\tilde{\boldsymbol{\xi}}_t)}{1+\beta} + (p_t-b)\tilde{z}_t - c_t\tilde{z}_{t-1} \\
& \qquad\qquad +(b-c_t)y_t(\tilde{\boldsymbol{\xi}}_{t-1}) + c_t y_{t-1}(\tilde{\boldsymbol{\xi}}_{t-2}) && \forall t \in \{2,...,T-1\} \\
& f_T(\tilde{\boldsymbol{\xi}}_T) \leq w_T(\tilde{\boldsymbol{\xi}}_{T-1}) + (p_T+h)\tilde{z}_T - c_T\tilde{z}_{T-1} \\
& \qquad\qquad -(h+c_T)y_T(\tilde{\boldsymbol{\xi}}_{T-1}) + c_T y_{T-1}(\tilde{\boldsymbol{\xi}}_{T-2}) \\
& f_T(\tilde{\boldsymbol{\xi}}_T) \leq w_T(\tilde{\boldsymbol{\xi}}_{T-1}) + (p_T-b)\tilde{z}_T - c_T\tilde{z}_{T-1} \\
& \qquad\qquad +(b-c_T)y_T(\tilde{\boldsymbol{\xi}}_{T-1}) + c_T y_{T-1}(\tilde{\boldsymbol{\xi}}_{T-2}).
\end{aligned}
\tag{34}
$$

We use the similar parameter setting as in Chen et al. (2007a). In particular, we set $h=6$, $b=3$, $\beta=0.1$, $x_1=w_1=0$, and $c_t=1$, $p_t=8$ for all $t\in[T]$. We set $\kappa\in\{0.25,1,4,16,64,256\}$ and assume each random demand $\tilde{z}_t$ is uniformly distributed over $\{0,1,2,...,20\}$.

We conduct two sets of experiments. In the first experiment, we assume that the decision-maker knows the exact demand distribution. We solve the problem using two approaches: DP and MLDR, where the expectation is taken with respect to the true distribution. Our goal is to show that MLDR is near-optimal as a policy by using DP as a benchmark since DP can obtain the optimal policy. After solving the corresponding problems, we implement the optimal policy obtained from DP and MLDR on 10,000 i.i.d. samples generated from the same underlying distribution. We average the results over 100 random instances. We report the out-of-sample multi-period CARA certainty equivalent of the consumption profile obtained from the two approaches under different risk tolerance parameter κ in Figure 2. We find that the multi-period MLDR policy performs comparably to the optimal DP policy across different risk tolerance levels.

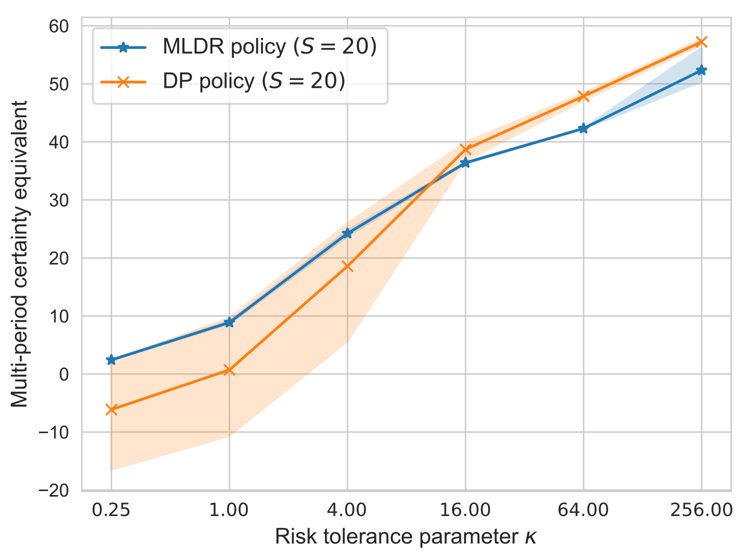
In the second experiment, we assume that the demand distribution is unknown, but the decision-maker has access to the empirical distribution of $S=20$ i.i.d. samples of $\tilde{\boldsymbol{z}}$ from the underlying distribution for each problem instance. Our goal is to evaluate the robustness of our MLDR policy in distribution ambiguity in a risk-averse setting. To achieve this, we solve the problem using DP, where the expectation is evaluated on the empirical distribution, and MLDR, where we maximize the multi-period ambiguity-averse CARA certainty equivalent $\mathbb{C}_{\mathcal{F}}^{\kappa,\boldsymbol{\theta}}\left[\boldsymbol{f}(\tilde{\boldsymbol{z}})\right]$ with an ambiguity set

Figure 2 Multi-period CARA certainty equivalent under different risk tolerance parameters (known distribution)



$\mathcal{F}$ based on mean, support, and MAD information estimated from the empirical distribution. We report the out-of-sample multi-period CARA certainty equivalent of the consumption profile in Figure 3 by visulizing the tube between the 10% and 90% quantiles (shaded areas) as well as the mean value (solid lines) over 100 random instances.

Figure 3 Multi-peirod CARA certainty equivalent under different risk tolerance parameters



Based on Figure 3, we observe that for high-risk tolerance levels ($\kappa \in \{16, 64, 256\}$), the DP policy outperforms the MLDR policy. However, for low-risk tolerance levels ($\kappa \in \{0.25, 1, 4\}$), the MLDR policy performs better than the DP policy, and we also observe that the multi-period CARA certainty equivalent under the MLDR policy is more concentrated, demonstrating the robustness of the MLDR-based approach. When compared with the perfect information setting in Figure 2, we find that under the data-driven setting, the DP policy deteriorates under high-risk aversion, while the MLDR policy can be advantageous by incorporating robustness.

It is important to note that this is a relatively simple multi-period model where we could obtain the optimal policy reasonably well using DP. The assumptions needed to obtain a tractable DP formulation can be quite fragile. For instance, if the borrowing and saving rates are different, the state space will significantly be enlarged and it may not be as computationally viable to solve for the optimal policy via DP. In contrast, we can easily incorporate these changes in our framework. The fact that the approximate MLDR policy performs reasonably well against the optimal DP policy is therefore a comforting assurance attesting to the effectiveness of the hierarchy of approximations that we have introduced to solve the multi-period robust CARA optimization problem.

References

Abbas AE, Howard RA (2015) *Foundations of Decision Analysis* (Pearson Higher Ed).

Allais M (1953) Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école américaine. *Econometrica: Journal of The Econometric Society* 503–546.

Anscombe FJ, Aumann RJ (1963) A definition of subjective probability. *Annals of Mathematical Statistics* 34(1):199–205.

Ardestani-Jaafari A, Delage E (2016) Robust optimization of sums of piecewise linear functions with application to inventory problems. *Operations Research* 64(2):474–494.

Arrow KJ (1965) Aspects of the theory of risk bearing .

Bajeux-Besnainou I, Portait R (1998) Dynamic asset allocation in a mean-variance framework. *Management Science* 44(11-part-2):S79–S95.

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

Ben-Tal A, El Housni O, Goyal V (2020) A tractable approach for designing piecewise affine policies in two-stage adjustable robust optimization. *Mathematical Programming* 182:57–102.

Ben-Tal A, Goryashko A, Guslitzer E, Nemirovski A (2004) Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* 99(2):351–376.

Bertsimas D, Doan XV, Natarajan K, Teo CP (2010) Models for minimax stochastic linear optimization problems with risk aversion. *Mathematics of Operations Research* 35(3):580–602.

Bertsimas D, Hertog Dd, Pauphilet J, Zhen J (2022) Robust convex optimization: A new perspective that unifies and extends. *Mathematical Programming* 1–42.

Bertsimas D, Popescu I (2002) On the relation between option and stock prices: a convex optimization approach. *Operations Research* 50(2):358–374.

Bertsimas D, Sim M, Zhang M (2019) Adaptive distributionally robust optimization. *Management Science* 65(2):604–618.

Birge JR, Louveaux F (2011) *Introduction to Stochastic Programming* (Springer Science & Business Media).

Bouakiz M, Sobel MJ (1992) Inventory control with an exponential utility criterion. *Operations Research* 40(3):603–608.

Chen L, He L, Zhou YH (2023) An exponential cone programming approach for managing electric vehicle charging. *Operations Research, Forthcoming* .

Chen W, Sim M (2009) Goal-driven optimization. *Operations Research* 57(2):342–357.

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

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

Chen X, Sim M, Sun P, Zhang J (2008) A linear decision-based approximation approach to stochastic programming. *Operations Research* 56(2):344–357.

Chen X, Zhang Y (2009) Uncertain linear programs: Extended affinely adjustable robust counterparts. *Operations Research* 57(6):1469–1482.

Chen Z, Sim M, Xiong P (2020) Robust stochastic optimization made easy with RSOME. *Management Science* 66(8):3329–3339.

Chicoisne R, Ordóñez F, Espinoza D (2018) Risk averse shortest paths: A computational study. *INFORMS Journal on Computing* 30(3):539–553.

Corner JL, Corner PD (1995) Characteristics of decisions in decision analysis practice. *Journal of the Operational Research Society* 46(3):304–314.

Dantzig GB (1955) Linear programming under uncertainty. *Management Science* 1(3-4):197–206.

Delage E, Iancu DA (2015) Robust multistage decision making. *INFORMS TutORials in Operations Research* 20–46.

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

Delquié P (2008) Interpretation of the risk tolerance coefficient in terms of maximum acceptable loss. *Decision Analysis* 5(1):5–9.

Dowson O, Morton DP, Pagnoncelli BK (2020) Multistage stochastic programs with the entropic risk measure.

Dyer M, Stougie L (2006) Computational complexity of stochastic programming problems. *Mathematical Programming* 106(3):423–432.

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

Esfahani PM, Kuhn D (2018) Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming* 171(1-2):115–166.

Feng Y, Xiao B (2008) A risk-sensitive model for managing perishable products. *Operations Research* 56(5):1305–1311.

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

Frederick S, Loewenstein G, O'donoghue T (2002) Time discounting and time preference: A critical review. *Journal of Economic Literature* 40(2):351–401.

Garstka SJ, Wets RJB (1974) On decision rules in stochastic programming. *Mathematical Programming* 7(1):117–143.

Georghiou A, Wiesemann W, Kuhn D (2015) Generalized decision rule approximations for stochastic programming via liftings. *Mathematical Programming* 152(1):301–338.

Gilboa I, Schmeidler D (1989) Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2):141–153.

Goh J, Sim M (2010) Distributionally robust optimization and its tractable approximations. *Operations Research* 58(4):902–917.

Grötschel M, Lovász L, Schrijver A (2012) *Geometric algorithms and combinatorial optimization*, volume 2 (Springer Science & Business Media).

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

Hanasusanto GA, Kuhn D, Wiesemann W (2016) A comment on "computational complexity of stochastic programming problems". *Mathematical Programming* 159(1-2):557–569.

Holmstrom B, Milgrom P (1991) Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *The Journal of Law, Economics, & Organization* 7:24.

Howard RA, Matheson JE (1972) Risk-sensitive Markov decision processes. *Management Science* 18(7):356–369.

Jaillet P, Qi J, Sim M (2016) Routing optimization under uncertainty. *Operations Research* 64(1):186–200.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47(2):363–391.

Kall P, Wallace SW, Kall P (1994) *Stochastic Programming* (Springer).

Kirkwood CW (2004) Approximating risk aversion in decision analysis applications. *Decision Analysis* 1(1):51–67.

Knight FH (1921) Risk, uncertainty and profit .

Kydland FE, Prescott EC (1977) Rules rather than discretion: The inconsistency of optimal plans. *Journal of Political Economy* 85(3):473–491.

Laibson D (1997) Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics* 112(2):443–478.

Loch CH, Wu Y (2007) *Behavioral Operations Management* (Now Publishers Inc).

Markowitz H (1952) Portfolio selection. *The Journal of Finance* 7(1):77–91.

Matsui T (1996) NP-hardness of linear multiplicative programming and related problems. *Journal of Global Optimization* 9:113–119.

Mehra R, Prescott EC (1985) The equity premium: A puzzle. *Journal of Monetary Economics* 15(2):145–161.

Natarajan K, Pachamanova D, Sim M (2008) Incorporating asymmetric distributional information in robust value-at-risk optimization. *Management Science* 54(3):573–585.

Nemirovski A, Shapiro A (2007) Convex approximations of chance constrained programs. *SIAM Journal on Optimization* 17(4):969–996.

Postek K, Ben-Tal A, Den Hertog D, Melenberg B (2018) Robust optimization with ambiguous stochastic constraints under mean and dispersion information. *Operations Research* 66(3):814–833.

Pratt JW (1964) Risk aversion in the small and in the large. *Econometrica* 32(1/2):122–136.

Prékopa A (2013) *Stochastic Programming*, volume 324 (Springer Science & Business Media).

Quiggin J (1982) A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3(4):323–343.

Scarf HE (1957) A min-max solution of an inventory problem. Technical report, RAND CORP SANTA MONICA CALIF.

See CT, Sim M (2010) Robust approximation to multiperiod inventory management. *Operations Research* 58(3):583–594.

Shapiro A, Dentcheva D, Ruszczyński A (2014) *Lectures on Stochastic Programming: Modeling and Theory* (SIAM).

Simon HA (1955) A behavioral model of rational choice. *The Quarterly Journal of Economics* 99–118.

Smith JE, Winkler RL (2006) The optimizer's curse: Skepticism and postdecision surprise in decision analysis. *Management Science* 52(3):311–322.

Toh KC (2018) Some numerical issues in the development of SDP algorithms. *INFORMS OS Today* 8(2):7–20.

Trefethen LN (2019) *Approximation Theory and Approximation Practice, Extended Edition* (SIAM).

Varian HR (1992) *Microeconomic Analysis*, volume 3 (Norton New York).

Veronesi P (1999) Stock market overreactions to bad news in good times: a rational expectations equilibrium model. *The Review of Financial Studies* 12(5):975–1007.

Von Neumann J, Morgenstern O (1947) *Theory of Games and Economic Behavior (commemorative edition)* (Princeton University Press).

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

Ye Q, Xie W (2021) Second-order conic and polyhedral approximations of the exponential cone: Application to mixed-integer exponential conic programs. *arXiv:2106.09123* .

Yu Q, Allon G, Bassamboo A (2017) How do delay announcements shape customer behavior? an empirical study. *Management Science* 63(1):1–20.

Zhang Y, Zhang Z, Lim A, Sim M (2021) Robust data-driven vehicle routing with time windows. *Operations Research* 69(2):469–485.

Zhen J, Den Hertog D, Sim M (2018) Adjustable robust optimization via Fourier–Motzkin elimination. *Operations Research* 66(4):1086–1100.

Appendix A: Proofs of results

Proof of Proposition 1. The proof is the same as that of Lemma 1 in Jaillet et al. (2016) and thus omitted. □

Proof of Proposition 2. We only prove the super-additivity as follows since the proof of another property can be referred to Lemma 1 in Jaillet et al. (2016). For any $\kappa_1, \kappa_2 > 0$, let $\kappa = \kappa_1 + \kappa_2$, we have

$$\mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_1 + \tilde{v}_2] = \mathbb{C}^{\kappa}_{\mathbb{P}}\left[\frac{\kappa_1}{\kappa}\frac{\kappa\tilde{v}_1}{\kappa_1} + \frac{\kappa_2}{\kappa}\frac{\kappa\tilde{v}_2}{\kappa_2}\right] \geq \frac{\kappa_1}{\kappa}\mathbb{C}^{\kappa}_{\mathbb{P}}\left[\frac{\kappa\tilde{v}_1}{\kappa_1}\right] + \frac{\kappa_2}{\kappa}\mathbb{C}^{\kappa}_{\mathbb{P}}\left[\frac{\kappa\tilde{v}_2}{\kappa_2}\right] = \mathbb{C}^{\kappa_1}_{\mathbb{P}}[\tilde{v}_1] + \mathbb{C}^{\kappa_2}_{\mathbb{P}}[\tilde{v}_2]$$

where the inequality is from concavity of CARA certainty equivalent in Proposition 1. For the cases of either κ_1 or κ_2 is zero, we assume $\kappa_1 = \kappa$ and $\kappa_2 = 0$ without loss of generality. Then by the last property in Proposition 1, we have

$$\mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_1 + \tilde{v}_2] = \mathbb{C}^{\kappa}_{\mathbb{P}}\left[\tilde{v}_1 + \tilde{v}_2 - \mathbb{C}^{0}_{\mathbb{P}}[\tilde{v}_2]\right] + \mathbb{C}^{0}_{\mathbb{P}}[\tilde{v}_2] \geq \mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_1] + \mathbb{C}^{0}_{\mathbb{P}}[\tilde{v}_2].$$

The super-additivity of $\mathbb{C}^{\kappa}_{\mathcal{F}}[\tilde{v}]$ can be proved in the same way. □

Proof of Theorem 2. We denote $\bar{\mathcal{U}} = \left\{(\boldsymbol{x}, \kappa, y) \,\middle|\, \exists\, \boldsymbol{q} \in \mathbb{R}^I :\ \sum_{i\in[I]} p_i q_i \leq \kappa, (q_i, \kappa, y - x_i) \in \mathcal{K}_{\exp}, \forall i \in [I]\right\}$ and $\mathcal{U} = \{(\boldsymbol{x}, \kappa, y) | y \leq g(\boldsymbol{x}, \kappa), \kappa > 0\}$. Observe that

$$\begin{aligned}\mathcal{U} &= \left\{(\boldsymbol{x}, \kappa, y) \,\middle|\, -\kappa \log \sum_{i\in[I]} p_i e^{-x_i/\kappa} \geq y, \kappa > 0\right\}\\ &= \left\{(\boldsymbol{x}, \kappa, y) \,\middle|\, \sum_{i\in[I]} p_i \kappa e^{(y-x_i)/\kappa} \leq \kappa, \kappa > 0\right\}\\ &= \left\{(\boldsymbol{x}, \kappa, y) \,\middle|\, \exists\, \boldsymbol{q} \in \mathbb{R}^I :\ \sum_{i\in[I]} p_i q_i \leq \kappa, \kappa > 0, (q_i, \kappa, y - x_i) \in \mathcal{K}_{\exp}, \forall i \in [I]\right\}.\end{aligned}$$

Clearly $\mathcal{U} \subseteq \bar{\mathcal{U}}$ and the latter is closed. Hence the closure $\mathrm{cl}(\mathcal{U}) \subseteq \bar{\mathcal{U}}$.

Next, we show $\bar{\mathcal{U}} \subseteq \mathrm{cl}(\mathcal{U})$. For any $(\boldsymbol{x}, \kappa, y) \in \bar{\mathcal{U}} \backslash \mathcal{U}$, we have $\kappa = 0$, $x_i \geq y$ for all $i \in [I]$. We denote $\bar{x} = \min_{i\in[I]}\{x_i\}$ and consider the sequence $\{(\boldsymbol{x}^j, \kappa^j, y^j)\}_{j=1}^{\infty} \in \mathcal{U}$ where $\boldsymbol{x}^j = \boldsymbol{x}$, $\kappa^j = 1/j$ and $y^j = \min\{y, g(\boldsymbol{x}^j, \kappa^j)\}$. Since

$$\lim_{j\to\infty} g(\boldsymbol{x}^j, \kappa^j) = \bar{x} - \lim_{j\to\infty} \kappa^j \log \sum_{i\in[I]} p_i e^{(\bar{x}-x_i)/\kappa^j} \leq \bar{x} - \lim_{j\to\infty} \kappa^j \log \sum_{\{i\in[I]: x_i = \bar{x}\}} p_i = \bar{x},$$

and

$$\bar{x} - \lim_{j\to\infty} \kappa^j \log \sum_{i\in[I]} p_i e^{(\bar{x}-x_i)/\kappa^j} \geq \bar{x} - \lim_{j\to\infty} \kappa^j \log \sum_{i\in[I]} p_i = \bar{x},$$

we know

$$\lim_{j\to\infty} y^j = \min\left\{y, g(\boldsymbol{x}^j, \kappa^j)\right\} = \min\{y, \bar{x}\} = y$$

and hence $\lim_{j\to\infty}(\boldsymbol{x}^j, \kappa^j, y^j) = (\boldsymbol{x}, \kappa, y)$. Therefore, $(\boldsymbol{x}, \kappa, y) \in \mathrm{cl}(\mathcal{U})$ and $\mathrm{cl}(\mathcal{U}) = \bar{\mathcal{U}}$. □

Proof of Theorem 1. Consider the following

Separation problem (S): Given $\kappa > 0$, t, $\boldsymbol{x}$, α, $\boldsymbol{\beta}$, check if the dual feasibility constraint (11) *is satisfied? If not, find a $\boldsymbol{z} \in \mathcal{Z}$ such that $\kappa \exp\left(\frac{t - \boldsymbol{x}^\top \boldsymbol{z}}{\kappa}\right) - \alpha - \boldsymbol{\beta}^\top \boldsymbol{z} > 0$.*

If this separation problem (S) is NP-hard. The equivalence of separation and optimization (Grötschel et al. 2012) then implies that the ambiguity-averse CARA certainty equivalent evaluation problem is NP-hard.

Inspired by Matsui (1996), we consider a reduction from the set partition problem:

Set partition problem (P): Given a 0-1 matrix $\boldsymbol{M} \in \{0,1\}^{m\times n}$ $(m<n)$*, is there a 0-1 valued solution* $\boldsymbol{x}\in\{0,1\}^n$ *to the system* $\boldsymbol{M}\boldsymbol{x}=\boldsymbol{1}$*?*

We first consider a quadratic optimization problem over a polytope,

$$\max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{Z}(\boldsymbol{M})} \left(\sum_{i\in[n]} p^i x_i\right)^2 - \sum_{i\in[n]}\sum_{j\in[n]} p^{i+j} y_{ij} \tag{35}$$

where $p\geq 2$ is an integer and the feasible region is

$$\mathcal{Z}(\boldsymbol{M}) \triangleq \left\{ (\boldsymbol{x},\boldsymbol{y})\in\mathbb{R}^{n+n^2}: \begin{array}{ll} 0\leq x_i\leq 1 & \forall i\in[n] \\ x_i = y_{ii} & \forall i\in[n] \\ 0\leq y_{ij}\leq 1 & \forall i\in[n], j\in[n]\backslash\{i\} \\ y_{ij}\leq x_i, y_{ij}\leq x_j, y_{ij}\geq x_i+x_j-1 & \forall i\in[n], j\in[n]\backslash\{i\} \\ \boldsymbol{M}\boldsymbol{x}=\boldsymbol{1} & \end{array} \right\}.$$

We denote by $v_1^*(\boldsymbol{M})$ the optimal value of problem (35). Then we have the following theorem.

THEOREM 9 **(Theorem 2.2 in Matsui (1996))**. *Let* $n>m$ *with* $n>5$ *and* $p=n^{n^4}$*, then the system* $\boldsymbol{M}\boldsymbol{x}=\boldsymbol{1}$ *has a 0-1 valued solution* $\boldsymbol{x}\in\{0,1\}^n$ *if and only if* $v_1^*(\boldsymbol{M})\geq 0$*. Moreover, if the 0-1 valued solution does not exist, then* $v_1^*(\boldsymbol{M})<-p$*.*

We now construct an instance of the separation problem (S) as follows,

$$\max_{(\boldsymbol{x},\boldsymbol{y})\in\mathcal{Z}(\boldsymbol{M})} \frac{\exp\left(\theta\sum_{i\in[n]} p^i x_i\right) - 1 - \theta\sum_{i\in[n]} p^i x_i - \sum_{i\in[n]}\sum_{j\in[n]} \theta^2 p^{i+j} y_{ij}/2}{\theta^2/2} \tag{36}$$

where we choose $\theta>0$ such that

$$\sup_{|t|\leq p^{n+1}} \left| \frac{\exp(\theta t) - 1 - \theta t - \theta^2 t^2/2}{\theta^2/2} \right| \leq 1/2.$$

Since $|\sum_{i\in[n]} p^i x_i| \leq \sum_{i\in[n]} p^i \leq p^{n+1}$ for any $\boldsymbol{x}\in[\boldsymbol{0},\boldsymbol{1}]$, we have

$$\left| \frac{\exp\left(\theta\sum_{i\in[n]} p^i x_i\right) - 1 - \theta\sum_{i\in[n]} p^i x_i}{\theta^2/2} - \left(\sum_{i\in[n]} p^i x_i\right)^2 \right| \leq 1/2, \quad \forall (\boldsymbol{x},\boldsymbol{y})\in\mathcal{Z}(\boldsymbol{M}).$$

We denote the optimal value of problem (36) as $v_2^*(\boldsymbol{M})$. It follows that

$$v_1^*(\boldsymbol{M}) + 1/2 \geq v_2^*(\boldsymbol{M}) \geq v_1^*(\boldsymbol{M}).$$

Then we can decide the answer to the set partition problem (P) by solving the problem (36). If $v_2^*(\boldsymbol{M})\geq 0$, then there exists some $\boldsymbol{x}\in\{0,1\}^n$ satisfying $\boldsymbol{M}\boldsymbol{x}=\boldsymbol{1}$. Otherwise, $v_2^*(\boldsymbol{M})<1/2-p$, which implies there is no $\boldsymbol{x}\in\{0,1\}^n$ satisfying $\boldsymbol{M}\boldsymbol{x}=\boldsymbol{1}$. Since the set partition problem (P) is NP-complete, we conclude that the separation problem (S) is NP-hard. □

Proof of Example 2. Note that $\phi(\kappa,\lambda) = \mathbb{C}^{\kappa}_{\mathcal{G}}[\lambda\tilde{z}] = -\kappa \log \sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}\left[e^{-\lambda\tilde{z}/\kappa}\right]$ and

$$
\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{G}} \mathbb{E}_{\mathbb{P}}[e^{-\lambda\tilde{z}}] \le & \inf_{\gamma\ge 0,\alpha,\beta} && \alpha+\beta\mu+\gamma\delta \\
& \quad\text{s.t.} && e^{-\lambda z} \le \alpha+\beta z+\gamma|z-\mu| \quad \forall z\in[-1,1] \\
= & \inf_{\gamma\ge 0,\alpha,\beta} && \alpha+\beta\mu+\gamma\delta \\
& \quad\text{s.t.} && e^{-\lambda z} \le \alpha+\beta z+\gamma(\mu-z) \quad \forall z\in[-1,\mu] \\
& && e^{-\lambda z} \le \alpha+\beta z+\gamma(z-\mu) \quad \forall z\in[\mu,1] \\
= & \inf_{\gamma\ge 0,\alpha,\beta} && \alpha+\beta\mu+\gamma\delta \\
& \quad\text{s.t.} && e^{\lambda} \le \alpha-\beta+\gamma(\mu+1) \\
& && e^{-\mu\lambda} \le \alpha+\beta\mu \\
& && e^{-\lambda} \le \alpha+\beta+\gamma(1-\mu) \\
= & \sup_{p_1,p_2,p_3\ge 0} && p_1 e^{\lambda}+p_2 e^{-\mu\lambda}+p_3 e^{-\lambda} \\
& \quad\text{s.t.} && p_1+p_2+p_3 = 1 \\
& && -p_1+\mu p_2+p_3 = \mu \\
& && (\mu+1)p_1+(1-\mu)p_3 \le \delta
\end{aligned}
$$

where the first inequality is by weak duality, the second equality is because the optimal solution of a convex maximization problem is attained at the boundary, and the third equality is due to linear optimization strong duality. Clearly, the worst-case distribution is attained by a three-point distribution with probability mass p_1, p_2, p_3 on $-1, \mu, 1$. Solving the last linear optimization problem in the above bound, we get $p_1 = \frac{\delta}{2(1+\mu)}$, $p_3 = \frac{\delta}{2(1-\mu)}$ and $p_2 = 1-p_1-p_3$ and conclude the proof. □

Proof of Theorem 3. We first note that

$$
\begin{aligned}
\mathbb{C}^{\kappa}_{\mathcal{F}}[f(\boldsymbol{x},\tilde{\boldsymbol{z}})] = & \sup_{\boldsymbol{\alpha},\boldsymbol{\beta}} \ \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\min_{i\in\mathcal{I}}\left\{\alpha_i+\boldsymbol{\beta}_i^\top\tilde{\boldsymbol{z}}\right\}\right] \\
& \ \text{s.t.} \ a_i^0(\boldsymbol{x}) \ge \alpha_i && \forall i\in\mathcal{I} \\
& \qquad a_i^j(\boldsymbol{x}) \ge \beta_i^j && \forall i\in\mathcal{I}, j\in\mathcal{J}^+ \\
& \qquad a_i^j(\boldsymbol{x}) \le \beta_i^j && \forall i\in\mathcal{I}, j\in\mathcal{J}^- \\
& \qquad a_i^j(\boldsymbol{x}) = \beta_i^j && \forall i\in\mathcal{I}, j\in\mathcal{J}.
\end{aligned}
\tag{37}
$$

Then for any $\boldsymbol{\gamma}\in\mathbb{R}^{I_z}$, we have

$$
\begin{aligned}
& \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\min_{i\in\mathcal{I}}\left\{\alpha_i+\boldsymbol{\beta}_i^\top\tilde{\boldsymbol{z}}\right\}\right] \\
= & \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}+\min_{i\in\mathcal{I}}\{\alpha_i+(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{z}}\}\right] \\
\ge & \sup_{\kappa_0+\kappa_1=\kappa,\boldsymbol{\kappa}\ge\mathbf{0}} \mathbb{C}^{\kappa_0}_{\mathcal{F}}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}\right]+\mathbb{C}^{\kappa_1}_{\mathcal{F}}\left[\min_{i\in\mathcal{I}}\{\alpha_i+(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{z}}\}\right] \\
= & \sup_{\kappa_0+\kappa_1=\kappa,\boldsymbol{\kappa}\ge\mathbf{0}} \mathbb{C}^{\kappa_0}_{\mathcal{F}}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}\right]-\kappa_1\log\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{\max_{i\in\mathcal{I}}\{-\alpha_i-(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{z}}\}}{\kappa_1}\right)\right] \\
\ge & \sup_{\kappa_0+\kappa_1=\kappa,\boldsymbol{\kappa}\ge\mathbf{0}} \mathbb{C}^{\kappa_0}_{\mathcal{F}}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}\right]-\kappa_1\log\sum_{i\in\mathcal{I}}\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{-\alpha_i+(\boldsymbol{\gamma}-\boldsymbol{\beta}_i)^\top\tilde{\boldsymbol{z}}}{\kappa_1}\right)\right] \\
= & \sup_{\kappa_0+\kappa_1=\kappa,\boldsymbol{\kappa}\ge\mathbf{0}} \mathbb{C}^{\kappa_0}_{\mathcal{F}}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}\right]-\kappa_1\log\sum_{i\in\mathcal{I}}\exp\left(-\frac{\mathbb{C}^{\kappa_1}_{\mathcal{F}}\left[\alpha_i+(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{z}}\right]}{\kappa_1}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \sup_{\boldsymbol{\kappa}\geq \mathbf{0},\boldsymbol{r}} r_0 - \kappa_1 \log \sum_{i\in\mathcal{I}} e^{-r_i/\kappa_1}\\
&\quad \text{s.t. } \kappa_0+\kappa_1=\kappa\\
&\qquad \mathbb{C}_{\mathcal{F}}^{\kappa_0}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}\right]\geq r_0\\
&\qquad \mathbb{C}_{\mathcal{F}}^{\kappa_1}\left[\alpha_i+(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{z}}\right]\geq r_i \qquad \forall i\in\mathcal{I}\\
&= \max_{\boldsymbol{\kappa}\geq\mathbf{0},\boldsymbol{r},\rho,\boldsymbol{q}} r_0+\rho\\
&\quad \text{s.t.}\quad \kappa_0+\kappa_1=\kappa\\
&\qquad \sum_{i\in\mathcal{I}} q_i\leq\kappa_1\\
&\qquad (q_i,\kappa_1,\rho-r_i)\in\mathcal{K}_{\exp} \qquad \forall i\in\mathcal{I}\\
&\qquad \sum_{j\in[I_z]}\phi_j(\kappa_0,\gamma^j)\geq r_0\\
&\qquad \alpha_i+\sum_{j\in[I_z]}\phi_j(\kappa_1,\beta_i^j-\gamma^j)\geq r_i \qquad \forall i\in\mathcal{I}
\end{aligned} \tag{38}
$$

where the first inequality is due to super-additivity of $\mathbb{C}_{\mathcal{F}}^{\kappa}[\tilde{v}]$ with respect to $(\kappa,\tilde{v})$ in Proposition 2, and the last equality is from Theorem 2. Combine (37) and (38) together and take infimum over all $\boldsymbol{\gamma}\in\mathbb{R}^{I_z}$, we obtain (14). □

Proof of Theorem 4. Without loss of generality, we can focus on proving the properties of the best lower bound (38) over $\boldsymbol{\gamma}\in\mathbb{R}^{I_z}$ of the ambiguity-averse CARA certainty equivalent $\mathbb{C}_{\mathcal{F}}^{\kappa}[\min_{i\in\mathcal{I}}\{\alpha_i+\boldsymbol{\beta}_i^\top\tilde{\boldsymbol{z}}\}]$ for any fixed $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Then all the conclusions in Theorem 4 can be obtained easily from the equivalence (37).

Since $\mathbb{C}_{\mathcal{F}}^{\kappa_0}[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}]$ is non-decreasing in κ_0, the lower bound (38) is non-decreasing in κ as one can fix κ_1 and increase κ_0 when κ becomes larger.

Consider the lower bound (38), note that $\kappa=0$ implies $\kappa_0=\kappa_1=0$, which further implies $\boldsymbol{q}=\mathbf{0}$ and $r_i\geq\rho$ for all $i\in\mathcal{I}$. Therefore, we must have

$$
\begin{aligned}
r_0 &= \sum_{j\in[I_z]}\phi_j(\kappa_0,\gamma^j)=\mathbb{C}_{\mathcal{F}}^{0}\left[\boldsymbol{\gamma}^\top\tilde{\boldsymbol{z}}\right]=\inf_{\boldsymbol{z}\in\mathcal{Z}}\boldsymbol{\gamma}^\top\boldsymbol{z}\\
r_i &= \alpha_i+\sum_{j\in[I_z]}\phi_j(\kappa_1,\beta_i^j-\gamma^j)=\mathbb{C}_{\mathcal{F}}^{0}\left[\alpha_i+(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\tilde{\boldsymbol{z}}\right]=\alpha_i+\inf_{\boldsymbol{z}\in\mathcal{Z}}(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\boldsymbol{z}\ \forall i\in\mathcal{I}\\
\rho &= \min_{i\in\mathcal{I}}\{r_i\}
\end{aligned}
$$

at optimality so that the best lower bound (38) over $\boldsymbol{\gamma}\in\mathbb{R}^{I_z}$ equals

$$
\sup_{\boldsymbol{\gamma}}\left(\inf_{\boldsymbol{z}\in\mathcal{Z}}\boldsymbol{\gamma}^\top\boldsymbol{z}+\min_{i\in\mathcal{I}}\left\{\alpha_i+\inf_{\boldsymbol{z}\in\mathcal{Z}}(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\boldsymbol{z}\right\}\right),
$$

which equals $\inf_{\boldsymbol{z}\in\mathcal{Z}}\min_{i\in\mathcal{I}}\{\alpha_i+\boldsymbol{\beta}_i^\top\boldsymbol{z}\}$ since

$$
\inf_{\boldsymbol{z}\in\mathcal{Z}}\min_{i\in\mathcal{I}}\left\{\alpha_i+\boldsymbol{\beta}_i^\top\boldsymbol{z}\right\}\geq\inf_{\boldsymbol{z}\in\mathcal{Z}}\boldsymbol{\gamma}^\top\boldsymbol{z}+\min_{i\in\mathcal{I}}\left\{\alpha_i+\inf_{\boldsymbol{z}\in\mathcal{Z}}(\boldsymbol{\beta}_i-\boldsymbol{\gamma})^\top\boldsymbol{z}\right\}
$$

for any $\boldsymbol{\gamma}\in\mathbb{R}^{I_z}$ and the equality holds when $\boldsymbol{\gamma}=\mathbf{0}$. Hence, the best lower bound over $\boldsymbol{\gamma}\in\mathbb{R}^{I_z}$ is exactly $\mathbb{C}_{\mathcal{F}}^{0}[\min_{i\in\mathcal{I}}\{\alpha_i+\boldsymbol{\beta}_i^\top\tilde{\boldsymbol{z}}\}]$.

If there is some $i^* \in \mathcal{I}$ such that $\alpha_{i^*} + \boldsymbol{\beta}_{i^*}^\top \boldsymbol{z} = \min_{i\in\mathcal{I}}\{\alpha_i + \boldsymbol{\beta}_i^\top \boldsymbol{z}\}$ for all $\boldsymbol{z} \in \mathcal{Z}$, then we let $\boldsymbol{\gamma} = \boldsymbol{\beta}_{i^*}$, $\kappa_0 = \kappa$ and $\kappa_1 = 0$ so that

$$
\begin{aligned}
r_0 &= \sum_{j\in[I_z]} \phi_j(\kappa, \gamma^j) = \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\boldsymbol{\gamma}^\top \tilde{\boldsymbol{z}}\right] = \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\boldsymbol{\beta}_{i^*}^\top \tilde{\boldsymbol{z}}\right] \\
r_i &= \alpha_i + \sum_{j\in[I_z]} \phi_j(\kappa_1, \beta_i^j - \gamma^j) = \mathbb{C}^{0}_{\mathcal{F}}\left[\alpha_i + (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i^*})^\top \tilde{\boldsymbol{z}}\right] = \alpha_i + \inf_{\boldsymbol{z}\in\mathcal{Z}} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i^*})^\top \boldsymbol{z} \;\; \forall i \in \mathcal{I} \\
\rho &= \min_{i\in\mathcal{I}} \{r_i\}
\end{aligned}
$$

at optimality and obtain the lower bound (38) as

$$
\begin{aligned}
r_0 + \rho &= \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\boldsymbol{\beta}_{i^*}^\top \tilde{\boldsymbol{z}}\right] + \min_{i\in\mathcal{I}}\left\{\alpha_i + \inf_{\boldsymbol{z}\in\mathcal{Z}} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i^*})^\top \boldsymbol{z}\right\} \\
&= \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\alpha_{i^*} + \boldsymbol{\beta}_{i^*}^\top \tilde{\boldsymbol{z}}\right] + \inf_{\boldsymbol{z}\in\mathcal{Z}} \min_{i\in\mathcal{I}} \left\{\alpha_i - \alpha_{i^*} + (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i^*})^\top \boldsymbol{z}\right\} \\
&\geq \mathbb{C}^{\kappa}_{\mathcal{F}}\left[\alpha_{i^*} + \boldsymbol{\beta}_{i^*}^\top \tilde{\boldsymbol{z}}\right],
\end{aligned}
$$

which implies the lower bound is greater than $\mathbb{C}^{\kappa}_{\mathcal{F}}\left[\min_{i\in\mathcal{I}}\{\alpha_i + \boldsymbol{\beta}_i^\top \tilde{\boldsymbol{z}}\}\right]$ and hence exact. □

Proof of Proposition 7. It follows from Jensen's inequality:

$$
\mathbb{E}_{\mathbb{P}^S}\left[-\kappa \log \frac{1}{S} \sum_{s\in[S]} \exp\left(-\frac{f(\boldsymbol{x}, \tilde{\boldsymbol{z}}^s)}{\kappa}\right)\right] \geq -\kappa \log \frac{1}{S} \sum_{s\in[S]} \mathbb{E}_{\mathbb{P}}\left[\exp\left(-\frac{f(\boldsymbol{x}, \tilde{\boldsymbol{z}}^s)}{\kappa}\right)\right] = \mathbb{C}^{\kappa}_{\mathbb{P}}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]
$$

as the function $-\kappa \log(\cdot)$ is convex for any $\kappa > 0$. □

Proof of Theorem 5. We claim that $\boldsymbol{c}^\top \hat{\boldsymbol{y}}(\boldsymbol{z}) \geq \boldsymbol{c}^\top \boldsymbol{y}^\dagger(\boldsymbol{z})$ for all $\boldsymbol{z} \in \mathcal{Z}$, which implies the conclusion directly. To show this, we note that for any $\boldsymbol{z} \in \mathcal{Z}$,

$$
\begin{aligned}
&\boldsymbol{c}^\top \hat{\boldsymbol{y}}(\boldsymbol{z}) - \boldsymbol{c}^\top \boldsymbol{y}^\dagger(\boldsymbol{z}) \\
&= \sum_{\ell\in[m]} \boldsymbol{c}^\top \boldsymbol{y}_*^\ell \left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|}\right\}\right)^+ - \sum_{i\in\mathcal{I}^o} \boldsymbol{c}^\top \boldsymbol{y}_\diamond^i \left(h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})\right)^+ \\
&= \sum_{\ell\in[m]} \boldsymbol{c}^\top \boldsymbol{y}_*^\ell \left(\left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|}\right\}\right)^+ - \sum_{i\in\mathcal{I}_\ell^o} \left(\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|}\right)^+\right).
\end{aligned}
$$

by noting $\boldsymbol{y}_\diamond^i = \boldsymbol{y}_*^\ell / \|\boldsymbol{b}_i\|$ for any $\ell \in [m]$, $i \in \mathcal{I}_\ell^o$. Hence it suffices to prove $\boldsymbol{c}^\top \boldsymbol{y}_*^\ell \leq 0$ for all $\ell \in [m]$ since

$$
\left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|}\right\}\right)^+ \leq \sum_{i\in\mathcal{I}_\ell^o} \max\left\{\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|}, 0\right\} = \sum_{i\in\mathcal{I}_\ell^o} \left(\frac{h_i(\boldsymbol{x}, \bar{\boldsymbol{y}}(\boldsymbol{z}), \boldsymbol{z})}{\|\boldsymbol{b}_i\|}\right)^+.
$$

Suppose there is some $\ell \in [m]$ such that $\boldsymbol{c}^\top \boldsymbol{y}_*^\ell > 0$ and $\boldsymbol{b}_i^\top \boldsymbol{y}_*^\ell \leq 0$ for all $i \in \mathcal{I}$, then for any $\boldsymbol{x}$ and $\boldsymbol{y}(\boldsymbol{z})$ feasible in Problem (17), the solution $\boldsymbol{y}(\boldsymbol{z}) + \lambda \boldsymbol{y}_*^\ell$ with any $\lambda > 0$ is also feasible. Hence the optimal value of Problem (17) is unbounded above, a contradiction. □

Proof of Theorem 6. Since the two-stage problem (16) has complete recourse with only one recourse decision variable, we must have $b_i > 0$ for all $i \in \mathcal{I}$ or $b_i < 0$ for all $i \in \mathcal{I}$. Observe that the second-stage linear optimization

$$
\begin{aligned}
f(\boldsymbol{x}, \boldsymbol{z}) = \max_{y} \;& b_0 y \\
\text{s.t.} \;& b_i y \leq a_i^0(\boldsymbol{x}) + \boldsymbol{a}_i^\top(\boldsymbol{x}) \boldsymbol{z} \quad \forall i \in \mathcal{I}
\end{aligned}
$$

is unbounded above if $b_0 b_i < 0$ for any $i \in \mathcal{I}$. Since the recourse decision y is unconstrained, for the optimal value of the problem to be finite, we can assume without loss of generality that $b_i > 0$ and $b_0 \geq 0$. In which case, the optimal decision rule $y^{OPT}(\boldsymbol{z}) = \min_{i\in\mathcal{I}}\left\{\frac{a_i^0(\boldsymbol{x})+\boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z}}{b_i}\right\}$ and $y_*^i = -1$ for all $i\in\mathcal{I}$. Hence, $m=1$ and $y_*^\ell = -1$ for all $\ell\in[1]$. Hence the MLDR is

$$\begin{aligned}\hat{y}(\boldsymbol{z}) &= y^0 + \boldsymbol{y}^\top\boldsymbol{z} - \left(\max_{i\in\mathcal{I}}\left\{\frac{b_i(y^0+\boldsymbol{y}^\top\boldsymbol{z}) - a_i^0(\boldsymbol{x}) - \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z}}{b_i}\right\}\right)^+\\ &= y^0 + \boldsymbol{y}^\top\boldsymbol{z} + \min\left\{0, \min_{i\in\mathcal{I}}\left\{-y^0 - \boldsymbol{y}^\top\boldsymbol{z} + \frac{a_i^0(\boldsymbol{x}) + \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z}}{b_i}\right\}\right\}\\ &= \min\left\{y^0+\boldsymbol{y}^\top\boldsymbol{z}, \min_{i\in\mathcal{I}}\left\{\frac{a_i^0(\boldsymbol{x})+\boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z}}{b_i}\right\}\right\}.\end{aligned}$$

Let $y^0 = a_i^0(\boldsymbol{x})/b_i$ and $\boldsymbol{y} = \boldsymbol{a}_i(\boldsymbol{x})/b_i$ for any $i\in\mathcal{I}$ we can recover $\hat{y}(\boldsymbol{z}) = y^{OPT}(\boldsymbol{z})$. □

Proof of Proposition 3. We prove it by case distinction.

(i) For $i\in\mathcal{I}\backslash\mathcal{I}^o$, we have $\boldsymbol{b}_i^\top\hat{\boldsymbol{y}}(\boldsymbol{z}) \le \boldsymbol{b}_i^\top\bar{\boldsymbol{y}}(\boldsymbol{z})$ for all $\boldsymbol{z}\in\mathcal{Z}$ since $\boldsymbol{b}_i^\top\boldsymbol{y}_*^\ell \le 0$ for all $\ell\in[m]$.

(ii) For $i\in\mathcal{I}^o$, let ℓ be the index such that $i\in\mathcal{I}_\ell^o$. For all $\boldsymbol{z}\in\mathcal{Z}$ we have

$$\begin{aligned}\boldsymbol{b}_i^\top\hat{\boldsymbol{y}}(\boldsymbol{z}) - a_i^0(\boldsymbol{x}) - \boldsymbol{a}_i^\top(\boldsymbol{x})\boldsymbol{z} &= h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z}) + \sum_{\ell\in[m]}\boldsymbol{b}_i^\top\boldsymbol{y}_*^\ell\left(\max_{j\in\mathcal{I}_\ell^o}\left\{\frac{h_j(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z})}{\|\boldsymbol{b}_j\|}\right\}\right)^+\\ &\le h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z}) + \boldsymbol{b}_i^\top\boldsymbol{y}_*^i\frac{(h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z}))^+}{\|\boldsymbol{b}_i\|}\\ &= \min\{h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z}),0\}\\ &\le 0\end{aligned}$$

where the first inequality is because $\left(\max_{j\in\mathcal{I}_\ell^o}\left\{\frac{h_j(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z})}{\|\boldsymbol{b}_j\|}\right\}\right)^+ \ge \frac{(h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z}))^+}{\|\boldsymbol{b}_i\|}$ and $\boldsymbol{b}_i^\top\boldsymbol{y}_*^k \le 0$ for all $k\in\mathcal{I}^o\backslash\{i\}$, and the second equality is due to $\boldsymbol{b}_i^\top\boldsymbol{y}_*^i = -\|\boldsymbol{b}_i\|$. □

Proof of Theorem 7. Note Problem (23) has the following lower bound:

$$\begin{aligned}&\mathbb{C}_\mathcal{F}^\kappa\left[\boldsymbol{c}^\top\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}) + \sum_{\ell\in[m]}\boldsymbol{c}^\top\boldsymbol{y}_*^\ell\left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}),\tilde{\boldsymbol{z}})}{\|\boldsymbol{b}_i\|}\right\}\right)^+\right]\\ &\ge \sup_{\substack{\kappa_0,\kappa_\ell\ge 0,\forall\ell\in[m]\\ \kappa_0+\sum_{\ell\in[m]}\kappa_\ell=\kappa}}\mathbb{C}_\mathcal{F}^{\kappa_0}\left[\boldsymbol{c}^\top\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}})\right] + \sum_{\ell\in[m]}\mathbb{C}_\mathcal{F}^{\kappa_\ell}\left[\boldsymbol{c}^\top\boldsymbol{y}_*^\ell\left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}),\tilde{\boldsymbol{z}})}{\|\boldsymbol{b}_i\|}\right\}\right)^+\right]\\ &= \sup_{\boldsymbol{\kappa}\ge\boldsymbol{0},\boldsymbol{r}} r_0 + \sum_{\ell\in[m]} r_\ell\\ &\quad \text{s.t.}\mathbb{C}_\mathcal{F}^{\kappa_0}\left[\boldsymbol{c}^\top\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}})\right]\ge r_0\\ &\quad\quad \mathbb{C}_\mathcal{F}^{\kappa_\ell}\left[\boldsymbol{c}^\top\boldsymbol{y}_*^\ell\left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}),\tilde{\boldsymbol{z}})}{\|\boldsymbol{b}_i\|}\right\}\right)^+\right]\ge r_\ell\\ &\quad\quad \kappa_0 + \sum_{\ell\in[m]}\kappa_\ell = \kappa\\ &= \sup_{\boldsymbol{\kappa}\ge\boldsymbol{0},\boldsymbol{r}} r_0 - \sum_{\ell\in[m]}(\boldsymbol{c}^\top\boldsymbol{y}_*^\ell)r_\ell\\ &\quad \text{s.t.}\mathbb{C}_\mathcal{F}^{\kappa_0}\left[\boldsymbol{c}^\top\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}})\right]\ge r_0\\ &\quad\quad \mathbb{C}_\mathcal{F}^{\kappa_\ell}\left[-\left(\max_{i\in\mathcal{I}_\ell^o}\left\{\frac{h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}),\tilde{\boldsymbol{z}})}{\|\boldsymbol{b}_i\|}\right\}\right)^+\right]\ge r_\ell\\ &\quad\quad \kappa_0 - \sum_{\ell\in[m]}(\boldsymbol{c}^\top\boldsymbol{y}_*^\ell)\kappa_\ell = \kappa\end{aligned} \tag{39}$$

where the first inequality is from super-additivity of $\mathbb{C}^{\kappa}_{\mathcal{F}}[\tilde{v}]$ in Proposition 2, and the last equality is because $\boldsymbol{c}^\top \boldsymbol{y}^{\ell}_* \leq 0$ for all $\ell \in [m]$ from Theorem 5. Combine (23) and (39), we obtain the following lower bound of Problem (17):

$$
\begin{aligned}
\max_{\substack{\boldsymbol{\kappa}\geq \boldsymbol{0},\\ \boldsymbol{r},\bar{\boldsymbol{y}}}} \quad & r_0 - \sum_{\ell\in[m]} (\boldsymbol{c}^\top \boldsymbol{y}^{\ell}_*) r_\ell && && \text{(40a)}\\
\text{s.t.} \quad & \kappa_0 - \sum_{\ell\in[m]} (\boldsymbol{c}^\top \boldsymbol{y}^{\ell}_*)\kappa_\ell = \kappa && && \text{(40b)}\\
& \mathbb{C}^{\kappa_0}_{\mathcal{F}}\left[\boldsymbol{c}^\top \bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}})\right] \geq r_0 && && \text{(40c)}\\
& \mathbb{C}^{\kappa_\ell}_{\mathcal{F}}\left[\min\left\{0, \min_{i\in\mathcal{I}^o_\ell}\left\{\frac{-h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\tilde{\boldsymbol{z}}),\tilde{\boldsymbol{z}})}{\|\boldsymbol{b}_i\|}\right\}\right\}\right] \geq r_\ell && \forall \ell\in[m] && \text{(40d)}\\
& h_i(\boldsymbol{x},\bar{\boldsymbol{y}}(\boldsymbol{z}),\boldsymbol{z}) \leq 0 && \forall \boldsymbol{z}\in\mathcal{Z}, \forall i\in\mathcal{I}\backslash\mathcal{I}^o && \text{(40e)}\\
& \bar{\boldsymbol{y}} \in \mathcal{L}^{I_z, I_y} && && \text{(40f)}
\end{aligned}
$$

To derive tractable reformulation of the above problem, we know constraint (40c) can be tractably reformulated by equation (9), constraints (40d) have safe $\mathcal{K}_{\text{exp}}$-representable approximations from Theorem 3, and constraints (40e) are robust linear constraints with box uncertainty set $\mathcal{Z}$, which can be easily reformulated as tractable linear constraints by standard robust optimization techniques. The resultant tractable model is exactly Problem (24). □

Proof of Proposition 4. For notation simplicity, let $g_{\boldsymbol{\theta}}(\boldsymbol{\nu},\kappa) = -\kappa \log\left(\sum_{t\in[T]} \theta_t \exp\left(-\frac{\nu_t}{\kappa}\right)\right)$ and note that $g_{\boldsymbol{\theta}}(\boldsymbol{\nu},\kappa)$ can be viewed as the CARA certainty equivalent of a random variable $\tilde{\nu}$ which realizes as ν_t with probability θ_t, $t\in[T]$. We first prove the variational representation (25). For any $\kappa>0$, as $g_{\boldsymbol{\theta}}(\boldsymbol{\nu},\kappa)$ is non-decreasing in ν_t, the maximum is attained at $\nu_t = \mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_t]$ for all $t\in[T]$. Hence the right hand side of (25) equals $-\kappa\log\left(\sum_{t\in[T]}\theta_t \exp(-\mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_t]/\kappa)\right) = \mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}]$. Moreover, since $\mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_t]$ is non-decreasing in $\kappa>0$ and $g_{\boldsymbol{\theta}}(\boldsymbol{\nu},\kappa)$ is non-decreasing in $\kappa>0$ and $\boldsymbol{\nu}\in\mathbb{R}^T$, we have $\mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}]$ is non-decreasing in $\kappa>0$ from representation (25) and the limit cases are exactly $\min_{t\in[T]:\theta_t>0}\{\operatorname{ess\,inf}_{\mathbb{P}}[\tilde{v}_t]\}$ at $\kappa=0$ and $\sum_{t\in[T]}\theta_t\mathbb{E}_{\mathbb{P}}[\tilde{v}_t]$ at $\kappa=\infty$ according to Proposition 1. $\mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}]$ is jointly concave in $(\tilde{\boldsymbol{v}},\kappa)$ with $\kappa>0$ because its hypograph

$$
\left\{(\tilde{\boldsymbol{v}},\kappa,\rho) \,\middle|\, \exists\, \boldsymbol{\nu}\in\mathbb{R}^T \;:\; \kappa>0,\ g_{\boldsymbol{\theta}}(\boldsymbol{\nu},\kappa)\geq\rho,\ \mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}_t]\geq \nu_t, \forall t\in[T]\right\}
$$

is convex thanks to concavity of $g_{\boldsymbol{\theta}}(\boldsymbol{\nu},\kappa)$ and $\mathbb{C}^{\kappa}_{\mathbb{P}}[\tilde{v}]$ in Proposition 1. Finally, for any $\nu\in\mathbb{R}$, the equality $\mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}+\nu\mathbf{1}] = \mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}]+\nu$ is straightforward to verify. □

Proof of Proposition 5. The proof of the first five properties is almost the same as that in Proposition 4 and hence omitted. The last property follows from the observation

$$
\inf_{\mathbb{P}\in\mathcal{F}} \mathbb{C}^{\kappa,\boldsymbol{\theta}}_{\mathbb{P}}[\tilde{\boldsymbol{v}}] = -\kappa\log\left(\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\sum_{t\in[T]}\theta_t\exp\left(-\frac{\tilde{v}_t}{\kappa}\right)\right]\right) \geq -\kappa\log\left(\sum_{t\in[T]}\theta_t\sup_{\mathbb{P}\in\mathcal{F}}\mathbb{E}_{\mathbb{P}}\left[\exp\left(-\frac{\tilde{v}_t}{\kappa}\right)\right]\right).
$$

□

Proof of Proposition 6. Clearly $\hat{\boldsymbol{y}}_t$ depends only on $\boldsymbol{\xi}_t$ and hence satisfies non-anticipativity. To show feasibility, note that for each $t\in[T]$, $i\in\mathcal{I}_t^o$, we have

$$\begin{aligned}
&\sum_{\tau\in[t]} \boldsymbol{b}_{t,i,\tau}^\top \hat{\boldsymbol{y}}_\tau(\boldsymbol{\xi}_\tau) - a_{t,i}^0(\boldsymbol{x}) - \boldsymbol{a}_{t,i}^\top(\boldsymbol{x})\boldsymbol{\xi}_t \\
&= h_{t,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t) + \sum_{\tau\in[t]} \boldsymbol{b}_{t,i,\tau}^\top \left(\sum_{s\in[\tau]}\sum_{\ell\in[m_s]} \boldsymbol{y}_{\tau*}^{s,\ell} \left(\max_{k\in\mathcal{I}_{s,\ell}^o} \left\{ \frac{h_{s,k}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[s]}(\boldsymbol{\xi}_s),\boldsymbol{\xi}_s)}{\|\boldsymbol{b}_{s,k,s}\|} \right\} \right)^+ \right) \\
&= h_{t,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t) + \sum_{s\in[t]}\sum_{\ell\in[m_s]} \left(\sum_{\tau=s}^{t} \boldsymbol{b}_{t,i,\tau}^\top \boldsymbol{y}_{\tau*}^{s,\ell} \right) \left(\max_{k\in\mathcal{I}_{s,\ell}^o} \left\{ \frac{h_{s,k}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[s]}(\boldsymbol{\xi}_s),\boldsymbol{\xi}_s)}{\|\boldsymbol{b}_{s,k,s}\|} \right\} \right)^+ . \\
&\le h_{t,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t) + \left(\boldsymbol{b}_{t,i,t}^\top \boldsymbol{y}_{t*}^{t,i}\right) \left(h_{t,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t)\right)^+ / \|\boldsymbol{b}_{t,i,t}\| \\
&= \min\left\{h_{t,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t), 0\right\} \\
&\le 0
\end{aligned}$$

where the first inequality is because $\sum_{\tau=s}^t \boldsymbol{b}_{t,i,\tau}^\top \boldsymbol{y}_{\tau*}^{s,\ell} \le 0$ for each $s\in[t]$ and $\ell\in[m_s]$ so that we can focus only on the case of $s=t$, ℓ such that $i\in\mathcal{I}_{t,\ell}^o$, and $k=i$, and the last equality is because $\boldsymbol{b}_{t,i,t}^\top \boldsymbol{y}_{t*}^{t,i} = -\|\boldsymbol{b}_{t,i,t}\|$. ☐

Proof of Theorem 8. We first note that Problem (27) is equivalent to:

$$\begin{aligned}
\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\nu}\in\mathbb{R}^T,\boldsymbol{y}_1,\ldots,\boldsymbol{y}_T} \quad & -\kappa\log\left(\sum_{t\in[T]} \theta_t e^{-\nu_t/\kappa}\right) \\
\text{s.t.} \quad & \mathbb{C}_{\mathcal{F}}^\kappa\left[\boldsymbol{c}_t^\top \boldsymbol{y}_t(\tilde{\boldsymbol{\xi}}_t)\right] \ge \nu_t && \forall t\in[T] \\
& h_{t,i}(\boldsymbol{x},\boldsymbol{y}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t) \le 0 && \forall t\in[T],\ \forall i\in\mathcal{I}_t,\ \forall \boldsymbol{z}\in\mathcal{Z} \\
& \boldsymbol{y}_t \in \mathcal{R}^{I_{\xi_t}, I_{y_t}} && \forall t\in[T].
\end{aligned} \tag{41}$$

from representation (26). Then we apply the MLDR (30) to obtain a lower bound of Problem (41):

$$\max_{\boldsymbol{x}\in\mathcal{X},\boldsymbol{\nu}\in\mathbb{R}^T,\bar{\boldsymbol{y}}_1,\ldots,\bar{\boldsymbol{y}}_T} -\kappa\log\left(\sum_{t\in[T]} \theta_t e^{-\nu_t/\kappa}\right) \tag{42a}$$

$$\text{s.t.} \quad \mathbb{C}_{\mathcal{F}}^\kappa\left[\boldsymbol{c}_t^\top \bar{\boldsymbol{y}}_t(\tilde{\boldsymbol{\xi}}_t) + \sum_{s\in[t]}\sum_{\ell\in[m_s]} \boldsymbol{c}_t^\top \boldsymbol{y}_{t*}^{s,\ell} \left(\max_{i\in\mathcal{I}_{s,\ell}^o}\left\{\frac{h_{s,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[s]}(\tilde{\boldsymbol{\xi}}_s),\tilde{\boldsymbol{\xi}}_s)}{\|\boldsymbol{b}_{s,i,s}\|}\right\}\right)^+\right] \ge \nu_t \quad \forall t\in[T] \tag{42b}$$

$$h_{t,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[t]}(\boldsymbol{\xi}_t),\boldsymbol{\xi}_t) \le 0 \qquad \forall t\in[T],\ \forall i\in\mathcal{I}_t\backslash\mathcal{I}_t^o,\ \forall \boldsymbol{z}\in\mathcal{Z} \tag{42c}$$

$$\bar{\boldsymbol{y}}_t \in \mathcal{L}^{I_{\xi_t}, I_{y_t}} \qquad \forall t\in[T]. \tag{42d}$$

according to Proposition 6. Note that the objective function (42a) is $\mathcal{K}_{\exp}$-representable by Theorem 2. Similar to the proof in Theorem 7, the constraints (42b) have safe approximations

$$\begin{aligned}
& r_0^t - \sum_{s\in[t]}\sum_{\ell\in[m_s]} (\boldsymbol{c}_t^\top \boldsymbol{y}_{t*}^{s,\ell}) r_{s,\ell}^t \ge \nu_t && \forall t\in[T] \\
& \kappa_0^t - \sum_{s\in[t]}\sum_{\ell\in[m_s]} (\boldsymbol{c}_t^\top \boldsymbol{y}_{t*}^{s,\ell}) \kappa_{s,\ell}^t = \kappa && \forall t\in[T] \\
& \mathbb{C}_{\mathcal{F}}^{\kappa_0^t}\left[\boldsymbol{c}_t^\top \bar{\boldsymbol{y}}_t(\tilde{\boldsymbol{\xi}}_t)\right] \ge r_0^t && \forall t\in[T] \\
& \mathbb{C}_{\mathcal{F}}^{\kappa_{s,\ell}^t}\left[\min\left\{0, \min_{i\in\mathcal{I}_{s,\ell}^o}\left\{-\frac{h_{s,i}(\boldsymbol{x},\bar{\boldsymbol{y}}_{[s]}(\tilde{\boldsymbol{\xi}}_s),\tilde{\boldsymbol{\xi}}_s)}{\|\boldsymbol{b}_{s,i,s}\|}\right\}\right\}\right] \ge r_{s,\ell}^t && \forall t\in[T], s\in[t], \ell\in[m_s],
\end{aligned}$$

in which the last constraints can be further safely approximated using Theorem 3 and the second-last constraints can be reformulated by equation (9). Finally, the robust linear constraints (42c) with box uncertainty set admit tractable robust counterparts as in Theorem 7. The resultant tractable model is exactly Problem (31). ☐

Appendix B: Exponential conic representations of ambiguous CARA certainty equivalent of payoff functions with affine perturbations

Table 1 **Equivalent representations of** $\phi(\kappa, \lambda)$

Ambiguity set	$\phi(\kappa,\lambda)$
$\mathcal{P}_0([-1,1])$	$-\lvert\lambda\rvert$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{P} \text{ is symmetric}\end{array}\right\}$	$-\kappa\log\left(\frac{e^{\lambda/\kappa}+e^{-\lambda/\kappa}}{2}\right)$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{P} \text{ is unimodal w.r.t. } 0\end{array}\right\}$	$-\kappa\log\int_0^1 e^{s\lvert\lambda\rvert/\kappa}\mathrm{d}s$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{P} \text{ is symmetric,}\\ \text{unimodal w.r.t. } 0\end{array}\right\}$	$-\lambda-\kappa\log\int_0^1 e^{-2\lambda s/\kappa}\mathrm{d}s$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{E}_{\mathbb{P}}[\tilde{z}]\in[\underline{\mu},\overline{\mu}]\end{array}\right\}$	$\min\left\{\begin{array}{l}-\kappa\log\left(\frac{(1+\underline{\mu})e^{-\lambda/\kappa}+(1-\underline{\mu})e^{\lambda/\kappa}}{2}\right),\\ -\kappa\log\left(\frac{(1+\overline{\mu})e^{-\lambda/\kappa}+(1-\overline{\mu})e^{\lambda/\kappa}}{2}\right)\end{array}\right\}$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{E}_{\mathbb{P}}[\tilde{z}]=\mu\\ \mathbb{E}_{\mathbb{P}}[\lvert\tilde{z}-\mu\rvert]\le\delta\end{array}\right\}$	$-\kappa\log\left(\frac{\delta}{2(\mu+1)}e^{\lambda/\kappa}+\frac{\delta}{2(1-\mu)}e^{-\lambda/\kappa}+\left(1-\frac{\delta}{2(\mu+1)}-\frac{\delta}{2(1-\mu)}\right)e^{-\mu\lambda/\kappa}\right)$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{E}_{\mathbb{P}}[\tilde{z}]=\mu\\ \mathbb{E}_{\mathbb{P}}[\lvert\tilde{z}\rvert^2]\le\sigma^2\end{array}\right\}$	$\min\left\{\begin{array}{l}-\kappa\log\left(\frac{(1-\mu)^2\exp\left(\frac{-(\mu-\sigma^2)\lambda}{(1-\mu)\kappa}\right)+(\sigma^2-\mu^2)\exp(-\lambda/\kappa)}{1-2\mu+\sigma^2}\right),\\ -\kappa\log\left(\frac{(1+\mu)^2\exp\left(\frac{-(\mu+\sigma^2)\lambda}{(1+\mu)\kappa}\right)+(\sigma^2-\mu^2)\exp(\lambda/\kappa)}{1+2\mu+\sigma^2}\right)\end{array}\right\}$
$\left\{\begin{array}{l}\mathbb{P}\in\mathcal{P}_0([-1,1]):\\ \mathbb{P} \text{ is symmetric,}\\ \mathbb{E}_{\mathbb{P}}[\lvert\tilde{z}\rvert^2]\le\sigma^2\end{array}\right\}$	$-\kappa\log\left(\frac{\sigma^2(e^{\lambda/\kappa}+e^{-\lambda/\kappa})}{2}+1-\sigma^2\right)$

We observe that each reformulation $\phi(\kappa,\lambda)$ presented in Table 1 is exactly $\mathcal{K}_{\exp}$-representable by Theorem 2, except for those involving integrals such as the following

$$\phi(\kappa,\lambda)=-\kappa\log\int_0^1 e^{s|\lambda|/\kappa}\mathrm{d}s=-\kappa\log\left(\frac{\kappa e^{|\lambda|/\kappa}-\kappa}{|\lambda|}\right).$$

Ben-Tal et al. (2009) derive its quadratic lower bound $-|\lambda|/2-\lambda^2/(24\kappa)$ based on Taylor's expansion. Nevertheless, we can use the Gaussian quadrature approximation (see, *e.g.,* Trefethen 2019)

$$\phi(\kappa,\lambda)=-\kappa\log\int_0^1 e^{s|\lambda|/\kappa}\mathrm{d}s\approx-\kappa\log\sum_{\ell\in[n]}\omega_\ell\exp\left(\frac{s_\ell|\lambda|}{\kappa}\right)$$

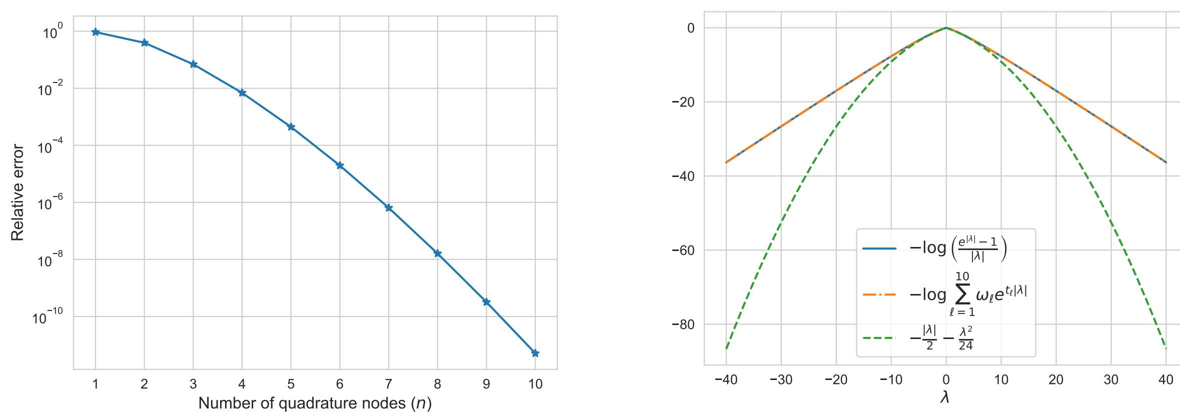
where ω_ℓ, s_ℓ, $\ell \in [n]$ are the quadrature weights and nodes on interval $[0,1]$. Clearly by Theorem 2 the quadrature approximation is $\mathcal{K}_{\exp}$-representable.

Motivated from the literature (*e.g.,* Yu et al. 2017), we can choose $n = 10$ with parameters

$$\begin{aligned} \boldsymbol{\omega} &= [0.0333, 0.0747, 0.1095, 0.1346, 0.1478, 0.1478, 0.1346, 0.1095, 0.0747, 0.0333] \\ \boldsymbol{s} &= [0.0130, 0.0675, 0.1603, 0.2833, 0.4256, 0.5744, 0.7167, 0.8397, 0.9325, 0.9870], \end{aligned}$$

which already provides very accurate estimation, see Figure 4(a) where we plot the relative approximation error of integral $\int_0^1 e^{10x}\mathrm{d}x$ by quadrature approximation with different number of nodes, and Figure 4(b) for a comparison of quadrature approximation and the quadratic lower bound of $-\log\left(\frac{e^{|\lambda|}-1}{|\lambda|}\right)$, where the relative error of quadrature approximation is less than 10^{-5}.

Figure 4 **Illustration of Gaussian quadrature approximation**



(a) Approximation of $\int_0^1 e^{10x}\mathrm{d}x$

(b) Approximation of $-\log\left(\frac{e^{|\lambda|}-1}{|\lambda|}\right)$

Appendix C: Illustrative Examples of the bound in Theorem 3

We next illustrate the bound in Theorem 3 through two examples.

Comparison with a bound in Nemirovski and Shapiro (2007)

First we note that Theorem 3 provides a new tractable lower bound for evaluating $\mathbb{E}_{\mathbb{P}}\left[\min\left\{\lambda^0 + \sum_{j\in[I_z]} \lambda^j \tilde{z}_j, 0\right\}\right]$, which is shown to be #P-hard in Example 1 under independent identically distributed (i.i.d.) uniform random factors. One well-known tractable lower bound (Nemirovski and Shapiro 2007, Chen et al. 2008) is based on the observation of $-\min\{-z, 0\} = (z)^+ \le \kappa \exp\left(\frac{z}{\kappa} - 1\right)$ for any $\kappa > 0$. We show in the following simple example that our approximation scheme may improve that in certain cases.

Example 4. Let $\tilde{z}$ be a standard normal random variable. Consider approximations of $\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}+\mu)^+\right] = \mu F(\mu) + \frac{e^{-\mu^2/2}}{\sqrt{2\pi}}$ where $F(\cdot)$ is the cumulative distribution function of $\tilde{z}$. The popular upper bound in Nemirovski and Shapiro (2007) is

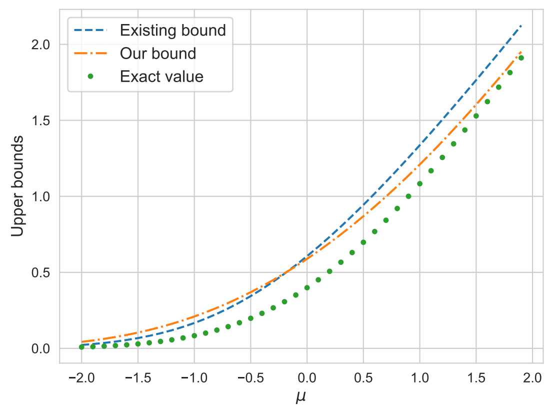
$$\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}+\mu)^+\right] \le \inf_{\kappa>0} \mathbb{E}_{\mathbb{P}}\left[\kappa \exp\left(\frac{\tilde{z}+\mu}{\kappa} - 1\right)\right] = \inf_{\kappa>0} \frac{\kappa}{e} \exp\left(\frac{\mu}{\kappa} + \frac{1}{2\kappa^2}\right), \tag{43}$$

while based on Theorem 3 we obtain the upper bound

$$\begin{aligned}\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}+\mu)^+\right] &= \overline{\mathbb{C}}_{\mathbb{P}}^{\infty}\left[(\tilde{z}+\mu)^+\right] = -\mathbb{C}_{\mathbb{P}}^{\infty}\left[\min\{-\tilde{z}-\mu, 0\}\right] \\ &\leq \inf_{\kappa_1>0,\gamma} \mathbb{E}_{\mathbb{P}}\left[\gamma\tilde{z}\right] + \kappa_1 \log \mathbb{E}_{\mathbb{P}}\left[\exp\left(\frac{(1-\gamma)\tilde{z}+\mu}{\kappa_1}\right) + \exp\left(\frac{-\gamma\tilde{z}}{\kappa_1}\right)\right] \\ &= \inf_{\kappa_1>0,\gamma} \kappa_1 \log\left(\exp\left(\frac{\mu}{\kappa_1} + \frac{(1-\gamma)^2}{2\kappa_1^2}\right) + \exp\left(\frac{\gamma^2}{2\kappa_1^2}\right)\right).\end{aligned} \tag{44}$$

See Figure 5 for a comparison of the existing bound (43) and our bound (44) where $\mu \in [-2, 2]$. We can see

Figure 5 **Upper bounds of** $\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}+\mu)^+\right]$



our bound is better than bound (43) when μ is mostly positive, though the bound does not perform as well when μ is mostly negative. We can improve the bound of $\mathbb{E}_{\mathbb{P}}\left[(\tilde{z}+\mu)^+\right]$ using infimal convolution of the two bounds (see, *e.g.*, Chen and Sim 2009). Unfortunately, for $\kappa < \infty$, we are not able to extend the bound (43) for evaluating the certainty equivalent, $\overline{\mathbb{C}}_{\mathbb{P}}^{\kappa}\left[(\tilde{z}+\mu)^+\right]$.

Comparison with the Monte-Carlo approximation

We next compare our bound with Monte-Carlo approximation, which is a typical way of evaluating the CARA certainty equivalent $\mathbb{C}_{\mathbb{P}}^{\kappa}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$ under distribution $\mathbb{P}$. Basically, we generate S i.i.d. samples from the distribution $\mathbb{P}$ and construct the random approximation of $\mathbb{C}_{\mathbb{P}}^{\kappa}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right]$ as follows:

$$-\kappa \log \frac{1}{S} \sum_{s\in[S]} \exp\left(-\frac{f(\boldsymbol{x}, \hat{\boldsymbol{z}}^s)}{\kappa}\right)$$

where $\hat{\boldsymbol{z}}^s$, $s \in [S]$ are realized samples independently drawn from $\tilde{\boldsymbol{z}} \sim \mathbb{P}$. We show the Monte-Carlo approximation is upward biased as follows.

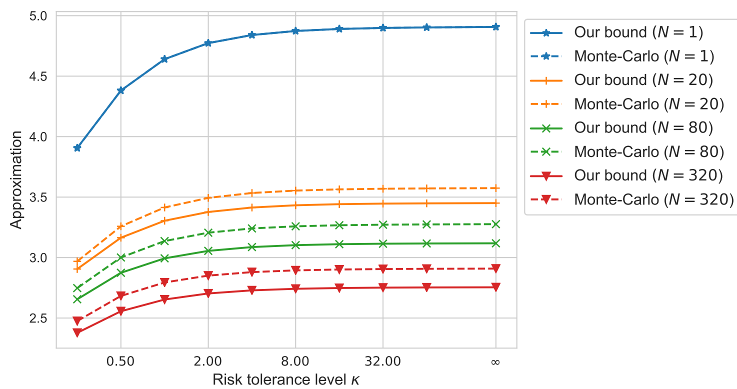
Proposition 7. *Consider the random vectors* $(\tilde{\boldsymbol{z}}^1, \ldots, \tilde{\boldsymbol{z}}^S) \sim \mathbb{P}^S$, *then*

$$\mathbb{E}_{\mathbb{P}^S}\left[-\kappa \log \frac{1}{S} \sum_{s\in[S]} \exp\left(-\frac{f(\boldsymbol{x}, \tilde{\boldsymbol{z}}^s)}{\kappa}\right)\right] \geq \mathbb{C}_{\mathbb{P}}^{\kappa}\left[f(\boldsymbol{x}, \tilde{\boldsymbol{z}})\right].$$

Although the Monte-Carlo approximation is upward biased, based on the law of large numbers, one can expect to obtain reasonably good approximation if the sample size S is large enough. This intuition holds true for high risk tolerance level. However, when risk tolerance level is low, we will show in the following example that the upward bias can be pronounced even with large sample size.

Example 5. We consider approximations of CARA certainty equivalent of the minimum of N weighted sum of $I_z = 20$ independently distributed random variables, for $N \in \{1, 20, 80, 320\}$. Specifically, we evaluate $\mathbb{C}_{\mathbb{P}}^{\kappa}\left[\min_{i\in[N]}\{\boldsymbol{a}_i^\top \tilde{\boldsymbol{z}}\}\right]$, where $\tilde{z}_j$, $j \in [I_z]$ are i.i.d. uniformly distributed random variables on $[0,1]$, and the weight vector $\boldsymbol{a}_i$ is randomly generated from the uniformly distributed unit hypercube $[0,1]^{I_z}$. We vary κ in $[0.25, 64]$ and include $\kappa = \infty$. The results are obtained by averaging over 50 random instances and presented in Figure 6. We see the Monte-Carlo approximation coincides with our lower bound when $N = 1$, suggesting

Figure 6 Comparison of our bound and Monte-Carlo approximation (10^6 samples) for $\kappa \geq 0.25$



both are accurate. For $N > 1$, the gap between the two approximations becomes larger as κ increases and stabilizes when $\kappa \to \infty$.

We observe that when the risk tolerance κ is low, the Monte-Carlo method may not provide accurate estimate; we show in Figure 7 where we plot the ratio of the Monte-Carlo approximations with $S \in \{10^4, 10^5, 10^6, 10^7\}$ samples to our bound of $\mathbb{C}_{\mathbb{P}}^{\kappa}[\boldsymbol{a}_1^\top \tilde{\boldsymbol{z}}]$ with $\kappa \in [0.05, 0.2]$. Note that our bound is exact at $N = 1$. We see the upward bias of Monte-Carlo approximation is more pronounced as S decreases, especially when $\kappa \leq 0.1$. What is surprising is the bias remains noticeable even with 10^7 samples. Hence, when the risk tolerance is low, the Monte-Carlo approximation would significantly overestimate the risk adjusted payoffs, while, as noted in Theorem 4, our deterministic approximation would provide a lower bound that is close to the actual CARA certainty equivalent. It is important to note that solving a stochastic optimization using SAA is a form of Monte-Carlo approximation, which yields random solutions with indicative objective values that are not achievable by the solutions. In contrast, our deterministic approximation would provide a pessimistic solution with an achievable indicative objective value.

Figure 7 **Ratio of Monte-Carlo approximation to our bound for $\kappa \leq 0.2$ at $N = 1$ where our bound is exact.**