PHBS WORKING PAPER SERIES

Control of Conservatism for Robust Optimization by Adjustable Regret

Yingjie Lan
Peking University

March 2023

Working Paper 20230306

Abstract

It is commonly recognized that a major issue of robust optimization is the tendency to produce overly conservative solutions. To address this issue, a new regret-based criterion with a single control parameter is proposed and axiomatized to offer smooth control of conservatism in a wide range without tampering with the uncertainty set. This criterion has many appealing analytical properties, such as decreasing conservatism with regard to the control parameter, which makes it a unique choice for fine control of conservatism. Tractability for robust linear programs with this criterion is established by reformulating them into those with the maximin criterion, for which tractable solution schemes and theoretical results are actively developed in the literature. Closed-form solutions are obtained for the robust one-way trading problem with this criterion, leading to a greatly simplified competitive ratio analysis. Numerical experiments are conducted to demonstrate smooth control of conservatism and the effects on revenue and risk.

Peking University HSBC Business School
University Town, Nanshan District
Shenzhen 518055, China

Continuous Control of Conservatism for Robust Optimization by Adjustable Regret

Yingjie Lan (email: ylan@pku.edu.cn)[a]

[a]*Peking University HSBC Business School, Xili University Town, Shenzhen, 518005, Guangdong, China*

Abstract

It is commonly recognized that a major issue of robust optimization is the tendency to produce overly conservative solutions. To address this issue, a new regret-based criterion with a single control parameter is proposed and axiomatized to offer smooth control of conservatism in a wide range without tampering with the uncertainty set. This criterion has many appealing analytical properties, such as decreasing conservatism with regard to the control parameter, which makes it a unique choice for fine control of conservatism. Tractability for robust linear programs with this criterion is established by reformulating them into those with the maximin criterion, for which tractable solution schemes and theoretical results are actively developed in the literature. Closed-form solutions are obtained for the robust one-way trading problem with this criterion, leading to a greatly simplified competitive ratio analysis. Numerical experiments are conducted to demonstrate smooth control of conservatism and the effects on revenue and risk.

Keywords: robust optimization, decision criteria, over-conservatism, tractability, one-way trading

1. Introduction.

Robust optimization (RO) in the broad sense deals with decision-making under data uncertainty without requiring an exact distribution, which is, in contrast, a prerequisite for stochastic programming. The survey of Gabrel et al. (2014) covers topics from distributional RO to the traditional or narrow-sensed RO that features an uncertainty set and a maximin reward or minimax cost criterion. RO is often considered more practical and powerful in applications where an exact distribution is difficult to estimate, due to little or inaccurate data, or nonstationarity of the underlying stochastic process. The traditional RO is initially developed in Soyster (1973), which proposes a robust counterpart for a linear programming (LP) model to address the feasibility concern, as data uncertainty can cause constraint violation. A large branch of studies follows Soyster (1973) to derive tractable reformulations that provide insights into robust solutions as well as probabilistic guarantees of feasibility (see Ben-Tal and Nemirovski 2008 and Bertsimas et al. 2011 for comprehensive surveys and Ben-Tal et al. 2009 for a book treatment).

As much performance is often sacrificed for robustness, it raises grave concern about optimality and over-conservatism, which is, for example, considered as a major reason preventing airlines from adopting robust revenue management methods in Vinod (2021). The robust counterpart of Soyster (1973) is quickly found to be overly conservative with a severe performance deterioration, which is caused by providing complete protection from all adverse scenarios. Such protection may be justifiable in engineering where infeasibility can cause a doomed satellite launch or a destroyed unmanned robot. However, adverse events in business like low demand or supply do not

bring about such disastrous consequences, thus protection can be made by a smaller and more flexible uncertainty set, with a lower probabilistic guarantee of feasibility in exchange for better performance. Significant progress in the theory of RO was then made in this regard for models with ellipsoidal uncertainty sets by Ben-Tal and Nemirovski (1998, 1999, 2000), El-Ghaoui and Lebret (1997), and El Ghaoui et al. (1998). As such models are nonlinear and much more demanding computationally than linear models, Bertsimas and Sim (2004) proposes the uncertainty budget method that can fully control the uncertainty set for every constraint while maintaining linearity. A drawback of these approaches is that they tamper with the uncertainty set to trade-off between robustness and conservatism, which can be difficult to balance, and may even be impossible in some engineering cases.

This paper tackles the issue of over-conservatism from a new angle without touching the uncertainty set by proposing a new decision criterion. Three criteria are commonly used in research and practices due to their good analytical properties and computational tractability, but none of them offers fine control of conservatism. The earliest maximin reward criterion is employed in traditional RO and often considered as the Achilles' heel, as it focuses entirely on worst-case profits and completely ignores all plausible opportunities for higher profits, which can lead to very conservative solutions. To alleviate this issue, the absolute and relative regret criteria are proposed on the basis of regret experienced by decision-makers once they realized what could have been the best action in hindsight (Kouvelis and Yu 2013). Both regret-based criteria are well-prepared to seize good opportunities with reduced level of conservatism, while sacrificing to some extent the assurance of worst-case

profit. The absolute regret criterion is proposed by Savage (1951), and axiomatized in Milnor (1951) and Stoye (2011). The relative regret criterion is equivalent to the so-called "competitive ratio," a popular measure for online optimization (Borodin and El-Yaniv 2005).

These criteria are empirically found to have distinct levels of conservatism in a consistent order. More and more studies describe regret-based criteria as less conservative than the maximin reward criterion (Perakis and Roels 2008, Natarajan et al. 2014, Wang et al. 2016, Caldentey et al. 2017, and Poursoltani and Delage 2021). In revenue management, Perakis and Roels (2008) observes that absolute regret decisions are less conservative than relative regret ones. This corroborates with the findings in numerical studies by Poursoltani and Delage (2021) that relative regret might be closer in spirit than absolute regret to maximin reward, which usually gives the most conservative solution.

This paper proposes adjustable regret minimization (ARM) as a new regret-based criterion capable of fine-tuning conservatism in a wide range without touching the uncertainty set. The new criterion is based on an interpolation between and extrapolation beyond the aforementioned criteria by a single control parameter. The parameter can be adjusted to suit the needs of an application, or adapted to key performance indicators (KPIs) monitored in real-time, such as in airline revenue management (Vinod 2021). It is intuitively explained in general and theoretically shown in a concrete example that as the control parameter increases, the level of conservatism of the ARM criterion decreases, which gives a clear direction for adjustment and adaptation.

4

Computational tractability is another important issue for RO applications. Averbakh and Lebedev (2005) shows that a robust LP with absolute regret criterion is strongly NP-hard. Since then, extensive efforts are made to develop exact and approximate solution schemes. Poursoltani and Delage (2021) provide general-purpose schemes by reformulating regret minimization problems into two-stage traditional RO problems, for which tractable solution schemes and theoretical knowledge are richly developed in the last decade (see Yanıkoğlu et al. 2019 for a recent survey). In this paper, parallel results are obtained for the ARM criterion, to maintain comparable tractability as the absolute regret criterion.

The ARM criterion also enables a new approach to competitive ratio analysis that can help reduce analysis complexity when closed-form solutions are available. Competitive ratio analysis is usually much more complex than absolute regret analysis, which can be seen, for example, by a contrast between El-Yaniv et al. (2001) and Wang et al. (2016), both of which solve the one-way trading problem under the same setting except for the criterion: competitive ratio for the former and absolute regret for the latter. This new approach for competitive ratio analysis may have a similar level of analysis complexity as that of the absolute regret analysis, as will be exemplified by solving the robust one-way trading problem with the ARM criterion. To facilitate such endeavors, the analytical properties of the ARM criterion are studied under a multistage setting, which readily covers single-stage and two-stage settings as special cases.

The main contributions of this paper are as follows: (1) The ARM criterion is proposed and axiomatized to offer fine control of conservatism in a

wide range for RO without touching the uncertainty set. It works with any specification of the uncertainty set, therefore can be applied either independently or jointly with methods based on modifying the uncertainty set, such as the uncertainty budget. (2) Linear RO problems with the ARM criterion are shown to have similar tractability as those with the absolute regret criterion. (3) The properties of the ARM criterion are studied theoretically to facilitate the analysis and its practical applications. Moreover, these theoretical properties are exploited to establish a new approach to competitive ratio analysis, which can reduce the complexity of analysis for some problems, especially when closed-form solutions are available. (4) The robust one-way trading problem with the ARM criterion is solved analytically, and the competitive ratio is derived by the new approach with reduced complexity than in the traditional way. Numerical experiments are conducted to demonstrate the effects of continuous control of conservatism.

The rest of this paper is arranged as follows. Section 2 gives general ARM-based formulations of multistage problems. Their properties are studied in Section 3 to facilitate analysis and applications, and the new approach to competitive ratio analysis is also discussed. Then section 4 deals with the tractability of linear problems, and in section 5 the robust one-way trading problem is studied to show the effectiveness of ARM. Finally, section 6 draws conclusions with future research suggestions.

2. Formulation.

The ARM criterion is first introduced for single-stage problems, which is then generalized to multistage problems. In a single-stage problem, an

action x is first taken from the robustly feasible set X, then a scenario ζ is realized from the uncertainty set $\mathcal{U}$, which may be continuous or discrete. The set X may also be continuous or discrete, for example, $X = \{x \in \mathcal{M} : \forall \zeta \in \mathcal{U}, g(x; \zeta) \leq 0\}$, where $g(x; \zeta) : X \times \mathcal{U} \to \mathbb{R}^{n_g}$, and $\mathcal{M} = \mathbb{R}^{n_r} \times \mathbb{Z}^{n_z}$ is a mixture of continuous and discrete space with $n_r, n_z \geq 0$ being the number of continuous and discrete components in x. The reward depends on both x and ζ, as specified by a reward function $r(x, \zeta)$. Let $r^*(\zeta) = \max_{x \in X} r(x, \zeta)$ denote the ex post optimal reward after having seen ζ being realized, which indicates the potential of a scenario. It is assumed that the min and max operators are well-defined, otherwise they could be replaced by inf and sup respectively.

A parameter $\beta \in [0, \infty)$ is introduced in ARM for continuous control of conservatism. The reward $r(x, \zeta)$ is compared with the β-adjusted benchmark $\beta r^*(\zeta)$ to obtain an adjustable regret $D(x, \zeta; \beta) = \beta r^*(\zeta) - r(x, \zeta)$. The worst-case regret $\bar{D}(x; \beta) = \max_{\zeta \in \mathcal{U}} D(x, \zeta; \beta)$ serves as a regret guarantee. The ARM criterion then chooses $x \in X$ with the best regret guarantee:

$$
\begin{aligned}
D(\beta) &= \min_{x \in X} \bar{D}(x; \beta) = \min_{x \in X} \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r(x, \zeta), \\
&= \min_{x \in X} \max_{\zeta \in \mathcal{U}} \beta \{ \max_{x' \in X} r(x', \zeta) \} - r(x, \zeta).
\end{aligned}
\tag{1}
$$

The ARM criterion unifies a few well-known robust criteria into a continuum as β takes on different values. At $\beta = 0$, it degenerates into the maximin reward criterion. With β at a special value between 0 and 1 (more on this later), it is equivalent to the relative regret criterion. Then at $\beta = 1$ it becomes the absolute regret criterion, and finally it transforms into the maximax criterion as $\beta \to \infty$. Note that as β gets bigger, the ARM criterion transmorphs into more aggressive criteria, which suggests that β can help

moderate the aggressiveness. Though the form of (1) was used for fractional combinatorial optimization in Megiddo (1978) and later adapted to numerically computing competitive ratios in Averbakh (2005), it has never been proposed as a new criterion for moderating conservatism as in this paper.

It is helpful to intuitively explain how such moderation happens as β increases. From the definition of $\bar{D}(x; \beta)$, there is $r(x, \zeta) \geq \beta r^*(\zeta) - D$ for any $\zeta \in \mathcal{U}$ if and only if $D \geq \bar{D}(x; \beta)$. Thinking of $r(x, \zeta)$ for a given x and $r^*(\zeta) - D$ as graphs over $\mathcal{U}$, clearly $r(x, \zeta) \geq \beta r^*(\zeta) - D$ says that the reward graph $r(x, \zeta)$ is entirely above the benchmark graph $\beta r^*(\zeta) - D$. As D decreases with the benchmark graph rising up, chances are that reward graphs no longer staying above are those that are either low in rewards or too dissimilar to the benchmark graph in that they have a low value while the benchmark has a high one at some ζ. When D reaches $D(\beta)$, the final reward graph that remains above and gets selected is likely to perform well and be similar to the benchmark graph. As β increases, the benchmark demands more rewards in favorable scenarios with higher $r^*(\zeta)$, and the recommended solution is likely to be more aggressive in grasping opportunities in favorable scenarios. The intuition will be further illustrated later in Corollary 3 for the one-way trading problem: the bigger the β, the more aggressive the recommended solution.

The single-stage setting can be readily extended to multistage, where decisions are made sequentially as the uncertainty gradually reveals itself stage by stage. Let $t = 1, \cdots, T$ labels the sequential stages, with a smaller t for an earlier stage. The decision variable x now consists of T subverctors $(x_1, \cdots, x_T)$, with the stage decision x_t corresponding to the decision in stage

t. Likewise, a whole scenario now consists of stage scenarios for each stage: $\zeta = (\zeta_1, \cdots, \zeta_T)$. Without loss of generality, the formulation can be standardized so that in each stage t the stage decision x_t first takes place, then the stage scenario ζ_t is realized afterward. In an application where a stage scenario is realized before a stage decision takes place, a dummy decision with only one choice of action (i.e. to start decision-making) can be inserted in the very beginning to have the standardized formulation, which helps simplify discussions, while the results are general nevertheless.

Just as in multistage stochastic programming (MSP), there is an implicit assumption: the realization of scenarios is independent of decisions, or the decision maker can not influence the scenario development. In the beginning of stage t, upon observing the partial scenario $\zeta_{1:t} = (\zeta_1, \cdots, \zeta_{t-1})$ revealed before stage t, the set of future scenarios includes only scenarios that share the same partially revealed scenario $\zeta_{1:t}$: $\mathcal{U}(\zeta_{1:t}) = \{\zeta' \in \mathcal{U} : \zeta'_{1:t} = \zeta_{1:t}\}$. It is convenient to define the stage scenario set $\mathcal{U}_t(\zeta_{1:t}) = \{\zeta'_t : \zeta' \in \mathcal{U}(\zeta_{1:t})\}$. Nonanticipativity as in MSP is held here by having the stage decision x_t dependent only on the stage decisions and stage scenarios in the earlier stages, so that decisions never depend on stage scenarios not revealed yet. Let $x_{1:t} = (x_1, \cdots, x_{t-1})$ be the partial sequence of stage decisions before stage t, and $h_t = (x_{1:t}, \zeta_{1:t})$ be the current history. Note that the set of robustly feasible actions in stage t depends not only on $x_{1:t}$, but also on $\zeta_{1:t}$ by reason of $\mathcal{U}(\zeta_{1:t})$. Therefore, let $X(h_t) = \{x' \in \mathcal{M} : x'_{1:t} = x_{1:t}, \forall \zeta' \in \mathcal{U}(\zeta_{1:t}), g(x'; \zeta') \leq 0\}$ denote the set of robustly feasible decisions and $X_t(h_t) = \{x_t : x \in X(h_t)\}$ denote all feasible actions in stage t.

The rewards may be accrued over the stages or may be received at once

in the end, let $r(x, \zeta)$ denote the total reward over all stages. Let $r^*(\zeta) = \max_{x \in X(\zeta)} r(x, \zeta)$ be the ex post optimal reward, where $X(\zeta) = \{x \in X | x_t \in X_t(x_{1:t}, \zeta_{1:t}), t = 1, \cdots, T\}$ is the set of all actions compatible with ζ. At the end of the last stage, the complete history $h_{T+1} = (x, \zeta)$ is known, and the regret is readily found by

$$D_T(h_{T+1}; \beta) = \beta r^*(\zeta) - r(x, \zeta). \tag{2}$$

Now it is possible to work from the last periods backwards to evaluate x_t in the context of h_t by this regret guarantee

$$\bar{D}_t(x_t, h_t; \beta) = \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t(h_{t+1}; \beta), \tag{3}$$

where h_{t+1} is formed by appending x_t and ζ_t to $x_{1:t}$ and $\zeta_{1:t}$ respectively. An optimal stage action x_t is chosen to minimize the regret guarantee

$$
\begin{aligned}
D_{t-1}(h_t; \beta) &= \min_{x_t \in X_t(h_t)} \bar{D}_t(x_t, h_t; \beta), \\
&= \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t(h_{t+1}; \beta). \tag{4}
\end{aligned}
$$

The definition (4) can be applied recursively for $t = T, \cdots, 1$ backwards, which gives a *plain formulation* with nonanticipativity. Note that when $t = 1$, there is no history in h_1, so let $D(\beta) = D_0(h_1; \beta)$, which is the best regret guarantee for the entire problem.

An alternative formulation is based on policies, where a policy π is a sequence of functions $\pi = \{\pi_t : h_t \to X_t(h_t), t = 1, 2, \cdots, T\}$ to make stage decisions according to $x_t = \pi_t(h_t)$, which takes care of nonanticipativity. Note that $X_t(h_t)$ can be replaced by the set of probabilistic mixtures of elements in $X_t(h_t)$ to allow for random policies, but the focus is on deterministic policies

10

for the sake of simplicity. The regret under a policy π is defined as follows. The regret with a full history $h_{T+1} = (x, \zeta)$ in the end of the last stage is simply

$$D_T^\pi(h_{T+1}; \beta) = \beta r^*(\zeta) - r(x, \zeta). \tag{5}$$

For $t = T, \cdots, 1$ the regret is defined recursively backwards by

$$D_{t-1}^\pi(h_t; \beta) = \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t^\pi(h_{t+1}^\pi; \beta), \tag{6}$$

where $h_{t+1}^\pi = ((x_{1:t}, \pi_t(h_t)), (\zeta_{1:t}, \zeta_t))$ denote the history evolution under π. To compute the overall regret $D^\pi(\beta) = D_0^\pi(h_1, \beta)$ (since h_1 is empty), simply apply (6) recursively to have

$$
\begin{aligned}
D^\pi(\beta) &= \max_{\zeta_1 \in \mathcal{U}_1(\zeta_{1:1})} D_1^\pi(h_2^\pi; \beta) \\
&= \max_{\zeta_1 \in \mathcal{U}_1(\zeta_{1:1})} \max_{\zeta_2 \in \mathcal{U}_2(\zeta_{1:2})} D_2^\pi(h_3^\pi; \beta) \\
&= \max_{\zeta_1 \in \mathcal{U}_1(\zeta_{1:1})} \cdots \max_{\zeta_T \in \mathcal{U}_T(\zeta_{1:T})} D_T^\pi(h_{T+1}^\pi; \beta) \\
&= \max_{\zeta \in \mathcal{U}} D_T^\pi(h_{T+1}^\pi; \beta) \tag{7}
\end{aligned}
$$

Let Π be the set of all policies, and $r^\pi(\zeta) = r(\pi(\zeta), \zeta)$, where $\pi(\zeta) = (\pi_1(h_1), \cdots, \pi_T(h_T))$ gives all stage decisions by policy π in scenario ζ. The policy-based formulation is given by

$$\min_{\pi \in \Pi} D^\pi(\beta) = \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^\pi(\zeta). \tag{8}$$

3. Properties.

The properties of the ARM criterion are studied in this section, to facilitate its application and the development of a new approach to competitive ratio analysis. The correspondence and equivalence between the two formulations is first established.

Theorem 1. *The plain formulation and the policy-based formulation have such a correspondence that for an arbitrary history h_t, there is*

$$D_{t-1}(h_t; \beta) = D_{t-1}^{\pi^*}(h_t; \beta), \ \textit{for } t = 1, \cdots, T+1, \tag{9}$$

with an optimal policy π^* constructed by

$$\pi_t^*(h_t) = \operatorname*{argmin}_{x_t \in X_t(h_t)} \max_{\zeta_t \in \mathcal{U}_t(h_t^\zeta)} D_t(h_{t+1}; \beta), t = 1, \cdots, T, \tag{10}$$

where the argmin *operator arbitrarily takes one minimizer when the solution is not unique.*

Proof: It is clear that (9) trivially holds for $t = T + 1$. For $t \leq T$, recall (4) and proceed as follows

$$
\begin{aligned}
D_{t-1}(h_t; \beta) &= \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t(h_{t+1}; \beta) \\
&= \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t^{\pi^*}(h_{t+1}^{\pi^*}; \beta) \\
&= D_{t-1}^{\pi^*}(h_t; \beta),
\end{aligned}
$$

where the second equality comes by (10), and the last equality comes by (6).

It remains to prove that π^* is optimal to (8) by showing for an arbitrary $\pi \in \Pi$ there is

$$D_{t-1}(h_t; \beta) \leq D_{t-1}^{\pi}(h_t; \beta), \tag{11}$$

for $t = 1, 2, \cdots, T + 1$ via backward induction on t. As the initial step, it trivially holds for $t = T + 1$. For the induction step, assume that (11) holds for $t + 1$: $D_t(h_{t+1}; \beta) \leq D_t^{\pi}(h_{t+1}; \beta)$, then show (11) also holds for t. Recall

(4) and replace $D_t(h_{t+1}; \beta)$ with $D_t^\pi(h_{t+1}; \beta)$ to have

$$
\begin{aligned}
D_{t-1}(h_t; \beta) &\leq \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t^\pi(h_{t+1}; \beta) \\
&\leq \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t^\pi(h_{t+1}^\pi; \beta) \\
&= D_{t-1}^\pi(h_t; \beta),
\end{aligned}
$$

where the second inequality comes by having $x_t = \pi_t(h_t)$, and the last equality comes by (6). Therefore (11) holds for all t by backward induction, and π^* is indeed an optimal policy. ∎

It is handy to have both formulations: the plain formulation is more useful in solving problems for practical applications, while the policy-based formulation can facilitate theoretical analysis. Clearly there is $D^{\pi^*}(\beta) = D(\beta)$ by (9) with $t = 1$. Besides this direct proof, an alternative proof of Theorem 1 can be made by recursively applying the interchangeability principle of Shapiro (2017), which involves much advanced mathematical concepts and is not adopted here. Conversely, the proof here can be adapted to make an alternative proof of the interchangeability principle. Note that the optimal policies given by (10) are "eager" as they always strive for the minimal regret (which can be smaller than $D(\beta)$ for some h_t), while there may be "lazy" optimal policies that deliver suboptimal objective values for such h_t (but still less than $D(\beta)$ to be optimal).

A decision criterion assigns a preference relation between any pair of policies π_1 and π_2. Policy π_1 is preferred to π_2 under the ARM criterion (denoted as $\pi_1 \succeq \pi_2$) if $D^{\pi_1}(\beta) \leq D^{\pi_2}(\beta)$, and strictly so (denoted as $\pi_1 \succ \pi_2$) if $D^{\pi_1}(\beta) < D^{\pi_2}(\beta)$.

Lemma 1. *Given a policy set Π, the ARM criterion holds for the following Axioms:*

Axiom 1 (complete ordering): *The relation satisfies (i) for any pair of policies π_1 and π_2, either $\pi_1 \preceq \pi_2$ or $\pi_2 \succeq \pi_1$; (ii) if $\pi_1 \succeq \pi_2 \succeq \pi_3$ holds for policies π_1, π_2, π_3, then $\pi_1 \succeq \pi_3$.*

Axiom 2 (symmetry): *The relation between any pair of policies is independent of the order in which the policies in Π are considered.*

Axiom 3 (strong domination): *If π_1 strongly dominates π_2 with $\forall \zeta \in \mathcal{U} : r(\pi_1(\zeta), \zeta) > r(\pi_2(\zeta), \zeta)$, then $\pi_1 \succ \pi_2$.*

Axiom 4 (continuity): *If a sequence of functions $\{r^i(;), i = 1, \cdots\}$ converges to the function $r(;)$ and if $r^i(\pi_1(\zeta), \zeta) \geq r^i(\pi_2(\zeta), \zeta)$ for all i, then $\pi_1 \succ \pi_2$.*

Axiom 5 (linearity): *The relation is not changed if $r(\pi(\zeta), \zeta)$ is replaced by $r'(\pi(\zeta), \zeta) \equiv \lambda r(\pi(\zeta), \zeta) + \mu$, and $\lambda > 0$.*

Axiom 6 (scenario randomization): *The relation is not changed if a new scenario ζ' is added, which is a randomized mixture of scenarios according to a distribution Υ on $\mathcal{U}$, with $r(\pi(\zeta'), \zeta') = E_{\zeta \sim \Upsilon}\{r(\pi(\zeta), \zeta)\}$.*

Axiom 7 (convexity): *If $\pi_1 \succeq \pi_2$ and $\pi_2 \succeq \pi_1$ (they are equally preferred), and $r(\pi(\zeta), \zeta) = (r(\pi_1(\zeta), \zeta) + r(\pi_2(\zeta), \zeta))/2$, then $\pi \succeq \pi_1$.*

Axiom 8 (special policy adjunction): *The ordering between the policies in Π is not changed by the adjunction of a new policy π', providing that $r(\pi'(\zeta), \zeta) \leq \max_{\pi \in \Pi} r(\pi(\zeta), \zeta)$ for all $\zeta \in \mathcal{U}$.*

Similar axioms as in Lemma 1 are also discussed in Milnor (1951) and Özkaya et al. (2022), and they are satisfied by both the maximin (Wald) and absolute regret (Savage) criterion, since the ARM criterion generalizes

both. The proof for these axioms is relatively simple. The complete ordering comes from the fact that $D^\pi(\beta)$ maps each policy π to a real number. Others can be proved by showing that the values of $D^\pi(\beta)$ for all $\pi \in \Pi$ are either unaffected or transformed by an order-preserving linear map.

Lemma 2. *The minimal regret guarantee $D_{t-1}(h_t; \beta)$ for all stages $t = 1, \cdots, T + 1$ with an arbitrary history h_t is continuous with regard to β.*

Proof: By backward induction on t from $T + 1$ to 1. When $t = T + 1$, it is clear that $D_T(h_{T+1}; \beta)$ is continuous in β according to (2), which completes the initial step. For the induction step, show that if $D_t(h_{t+1}; \beta)$ is continuous in β, then so is $D_{t-1}(h_t; \beta)$. It is clear that $\bar{D}_t(x_t, h_t; \beta)$ is continuous in β as it is a point-wise max of continuous functions by (3). Likewise, $D_{t-1}(h_t; \beta)$ is also continuous with regard to β by (4), which completes the proof. ∎

Theorem 2. *For $0 \le \beta_1 < \beta_2$, let $\pi_i^*, i \in \{1, 2\}$ be an optimal policy when $\beta = \beta_i$, and $\zeta_{ij}^* = \operatorname{argmax}_{\zeta \in \mathcal{U}} \beta_i r^*(\zeta) - r^{\pi_j^*}(\zeta), i, j \in \{1, 2\}$, then there is*

$$r^*(\zeta_{21}^*) \ge \frac{D(\beta_2) - D(\beta_1)}{\beta_2 - \beta_1} \ge r^*(\zeta_{12}^*). \tag{12}$$

Proof: By the definition of π_2^* and ζ_{12}^*, as well as Theorem 1, there is

$$
\begin{aligned}
D(\beta_1) &= \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta_1 r^*(\zeta) - r^\pi(\zeta) \\
&\le \max_{\zeta \in \mathcal{U}} \beta_1 r^*(\zeta) - r^{\pi_2^*}(\zeta) \\
&= \beta_1 r^*(\zeta_{12}^*) - r^{\pi_2^*}(\zeta_{12}^*).
\end{aligned}
$$

And there is $D(\beta_2) = \max_{\zeta \in \mathcal{U}} \beta_2 r^*(\zeta) - r^{\pi_2^*}(\zeta) \ge \beta_2 r^*(\zeta_{12}^*) - r^{\pi_2^*}(\zeta_{12}^*)$. There-

15

fore $D(\beta_2) - D(\beta_1) \geq (\beta_2 - \beta_1) \, r^*(\zeta_{12}^*)$. Similarly,

$$
\begin{aligned}
D(\beta_2) &= \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta_2 r^*(\zeta) - r^{\pi}(\zeta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta_2 r^*(\zeta) - r^{\pi_1^*}(\zeta) \\
&= \beta_2 r^*(\zeta_{21}^*) - r^{\pi_1^*}(\zeta_{21}^*).
\end{aligned}
$$

And there is $D(\beta_1) = \max_{\zeta \in \mathcal{U}} \beta_1 r^*(\zeta) - r^{\pi_1^*}(\zeta) \geq \beta_1 r^*(\zeta_{21}^*) - r^{\pi_1^*}(\zeta_{21}^*)$. Thus $D(\beta_2) - D(\beta_1) \leq (\beta_2 - \beta_1) \, r^*(\zeta_{21}^*)$. Therefore, (12) follows immediately. ∎

The continuity of Lemma 2 is a basic property useful for other analytical results. Note that when $\forall \zeta, r^*(\zeta) > 0$, Theorem 2 ensures the monotonicity of $D(\beta)$.

3.1. Convexity.

The convexity of $D(\beta)$ requires certain conditions to hold. Note that $\beta r^*(\zeta) - r^{\pi}(\zeta)$ is linear in β, thus the function

$$
F(\beta; \pi) = \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^{\pi}(\zeta)
$$

is convex in β for a given policy π. However, generally speaking, $D(\beta) = \min_{\pi \in \Pi} F(\beta; \pi)$ is not convex in β. In order for $D(\beta)$ to be convex, certain conditions are needed. A weak condition for convexity is introduced first.

Lemma 3. *A continuous function $f(y)$ on a convex domain Y is convex if*

$$
\forall y_1, y_2 \in Y \ \exists \lambda \in (0,1) \ f(\lambda y_1 + (1-\lambda)y_2) \leq \lambda f(y_1) + (1-\lambda)f(y_2). \quad (13)
$$

Proof: By contradiction. Assume $f(y)$ is not convex, then there exists $y_1, y_2 \in Y$ and $\lambda \in (0,1)$ such that $g(\lambda) > 0$, where $g(k) = f(y(k)) - (kf(y_1) + (1-k)f(y_2))$ and $y(k) = ky_1 + (1-k)y_2$. As $g(k)$ is continuous

16

with $g(0) = g(1) = 0$, there exists $k_1 = \max\{k \in [0, \lambda) : g(k) = 0\}, k_2 = \min\{k \in (\lambda, 1] : g(k) = 0\}$, such that $0 \le k_1 < \lambda < k_2 \le 1$, $g(k_1) = g(k_2) = 0$ and $\forall k' \in (k_1, k_2) \; g(k') > 0$.

Let $y'_1 = y(k_1), y'_2 = y(k_2)$, and $k' = \lambda' k_1 + (1 - \lambda') k_2$ for a $\lambda' \in (0, 1)$, then $y(k') = \lambda' y'_1 + (1 - \lambda') y'_2$. As $f(y'_i) = k_i f(y_1) + (1 - k_i) f(y_2), i = 1, 2$ from $g(k_1) = g(k_2) = 0$, there is $\lambda' f(y'_1) + (1 - \lambda') f(y'_2) = k' f(y_1) + (1 - k') f(y_2)$. By $g(k') > 0$ there is $f(y(k')) > k' f(y_1) + (1 - k') f(y_2)$, which implies $f(\lambda' y'_1 + (1 - \lambda') y'_2) > \lambda' f(y'_1) + (1 - \lambda') f(y'_2)$ for any $\lambda' \in (0, 1)$, which contradicts (13). ∎

A policy π dominates another policy π' (denoted as $\pi \ddot{\succeq} \pi'$) if for all $\zeta \in \mathcal{U}$ there is $r^\pi(\zeta) \ge r^{\pi'}(\zeta)$. Similarly, a scenario ζ dominates another scenario ζ' (denoted as $\zeta \ddot{\succeq} \zeta'$) if for all $\pi \in \Pi$ there is $r^\pi(\zeta) \le r^\pi(\zeta')$. According to Lan et al. (2008), dominated policies and scenarios can be eliminated by iterated elimination of dominated strategies in game theory.

Definition 1 (Reward Convexity (RC)). *The set Π has the property of RC if there is*

$$\forall \pi_1, \pi_2 \in \Pi \; \exists \pi \in \Pi \; \exists \lambda \in (0, 1) \; \forall \zeta \in \mathcal{U} \; r^\pi(\zeta) = \lambda r^{\pi_1}(\zeta) + (1 - \lambda) r^{\pi_2}(\zeta). \quad (14)$$

An example with the RC property is when all randomized policies (which randomly draw a deterministic policy from a probability distribution) are allowed and the reward of the random policy is given by the expected reward.

Definition 2 (Dominance Convexity (DC)). *The set Π has the DC property if there is*

$$\forall \pi_1, \pi_2 \in \Pi \; \exists \pi \in \Pi \; \exists \lambda \in (0, 1) \; \forall \zeta \in \mathcal{U} \; r^\pi(\zeta) \ge \lambda r^{\pi_1}(\zeta) + (1 - \lambda) r^{\pi_2}(\zeta). \quad (15)$$

Clearly, if Π has the RC property, then it also has the DC property. A more sophisticated example is as follows. If $r(x, \zeta)$ is concave in x and $X(\zeta)$ is convex for all $\zeta \in \mathcal{U}$, then (15) is satisfied. To see this, simply let $\pi(\zeta) = \pi_1(\zeta)/2 + \pi_2(\zeta)/2$. By the concavity of $r(x, \zeta)$ in x, there is $r(\pi(\zeta), \zeta) \geq r(\pi_1(\zeta), \zeta)/2 + r(\pi_2(\zeta), \zeta)/2$, and hence (15) is satisfied with $\lambda = 1/2$.

Both the RC and DC property are intact after scenario elimination. However, the RC property can be lost in policy elimination, the DC property still remains. Let $\hat{\Pi}$ be the set of all non-dominated policies in Π, so that any $\pi \in \Pi$ is dominated by a $\hat{\pi} \in \hat{\Pi}$. Let $\Pi^*(\beta)$ be the set of all optimal policies for a given β.

Theorem 3. *The DC property can transfer among $\Pi, \hat{\Pi}$ and $\Pi^*(\beta)$ (i) from Π to $\hat{\Pi}$, and vice versa; (ii) from Π to $\Pi^*(\beta)$, but not backward.*

Proof: (i.a) from Π to $\Pi^*(\beta)$. Let $\pi_1, \pi_2 \in \hat{\Pi} \subseteq \Pi$, thus there exists $\pi \in \Pi$ and $\lambda \in (0, 1)$ such that $r^{\pi}(\zeta) \geq \lambda r^{\pi_1}(\zeta) + (1 - \lambda)r^{\pi_2}(\zeta)$. As there is a $\hat{\pi} \in \hat{\Pi}$ such that $\hat{\pi} \ddot{\succeq} \pi$, it follows that $\hat{\Pi}$ has the DC property. (i.b) To show vice versa, let $\pi_1, \pi_2 \in \Pi$. Clearly there are $\hat{\pi}_1, \hat{\pi}_2 \in \hat{\Pi}$ such that $\hat{\pi}_1 \ddot{\succeq} \pi_1, \hat{\pi}_2 \ddot{\succeq} \pi_2$. There exists $\hat{\pi} \in \hat{\Pi}$ and $\lambda \in (0, 1)$ such that $r^{\hat{\pi}}(\zeta) \geq \lambda r^{\hat{\pi}_1}(\zeta) + (1 - \lambda)r^{\hat{\pi}_2}(\zeta) \geq \lambda r^{\pi_1}(\zeta) + (1 - \lambda)r^{\pi_2}(\zeta)$. As $\hat{\pi} \in \Pi$, hence Π also has the DC property.

(ii) from Π to $\Pi^*(\beta)$. Consider $\forall \pi_1^*, \pi_2^* \in \Pi^*(\beta) \subseteq \Pi$, by (15) there is a $\pi' \in \Pi$ such that

$$\exists \lambda \in (0, 1) \ \forall \zeta \in \mathcal{U} \ r^{\pi'}(\zeta) \geq \lambda r^{\pi_1^*}(\zeta) + (1 - \lambda)r^{\pi_2^*}(\zeta).$$

Now show that $\pi' \in \Pi^*(\beta)$ as follows:

$$
\begin{aligned}
D(\beta) &= \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^\pi(\zeta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^{\pi'}(\zeta) = D^{\pi'}(\beta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - \left(\lambda r^{\pi_1^*}(\zeta) + (1-\lambda) r^{\pi_2^*}(\zeta) \right) \\
&\leq \lambda \left(\max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^{\pi_1^*}(\zeta) \right) + \\
&\qquad (1-\lambda) \left(\max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^{\pi_2^*}(\zeta) \right) \\
&= \lambda D(\beta) + (1-\lambda)D(\beta) = D(\beta).
\end{aligned}
$$

Therefore $D^{\pi'}(\beta) = D(\beta)$ and so $\pi' \in \Pi^*(\beta)$. If $\Pi^*(\beta)$ is a singleton, then it has the DC property, but Π may not, which shows it does not transfer backward. ∎

Theorem 3 is useful to prove DC property for Π by simply focusing on the non-dominated subset $\hat{\Pi}$. Meanwhile, the DC property of $\Pi^*(\beta)$ may help select a more preferable policy when there are many optimal policies.

Theorem 4. *If Π has the DC property, then $D(\beta)$ is convex in β.*

Proof: Let π_i^* be an optimal policy for $\beta_i, i = 1, 2$. By (15) there exists $\pi' \in \Pi$ such that $\exists \lambda \in (0,1) \ \forall \zeta \in \mathcal{U} \ r^{\pi'}(\zeta) \geq \lambda r^{\pi_1^*}(\zeta) + (1-\lambda) r^{\pi_2^*}(\zeta)$. Let

$\beta = \lambda\beta_1 + (1 - \lambda)\beta_2$ and proceed as follows:

$$
\begin{aligned}
D(\beta) &= \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^\pi(\zeta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r^{\pi'}(\zeta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - \left(\lambda r^{\pi_1^*}(\zeta) + (1 - \lambda) r^{\pi_2^*}(\zeta) \right) \\
&\leq \lambda \left(\max_{\zeta \in \mathcal{U}} \beta_1 r^*(\zeta) - r^{\pi_1^*}(\zeta) \right) + \\
&\quad (1 - \lambda) \left(\max_{\zeta \in \mathcal{U}} \beta_2 r^*(\zeta) - r^{\pi_2^*}(\zeta) \right) \\
&= \lambda D(\beta_1) + (1 - \lambda) D(\beta_2).
\end{aligned}
$$

Therefore $D(\beta)$ is convex in β by Lemma 3. ∎

The DC property is relatively common, such as in most linear robust optimization problems, hence Theorem 4 readily applies. The convexity of $D(\beta)$ may help design faster algorithms than binary searches in Averbakh (2005) that numerically compute competitive ratios.

3.2. Competitive Ratio.

For reward maximization problems, the competitive ratio can be defined as

$$
\max_{\pi \in \Pi} \min_{\zeta \in \mathcal{U}} r_\zeta(\pi)/r_\zeta^*, \tag{16}
$$

which generally assumes $\forall \zeta \in \mathcal{U}, \ r^*(\zeta) > 0$. The relative regret criterion is recovered when β is set to the competitive ratio, as shown by the next lemma.

Lemma 4. *Assume $r^*(\zeta) > 0$ for all $\zeta \in \mathcal{U}$, the β_0 that solves $D(\beta) = 0$ is exactly the competitive ratio, and the set of optimal policies for (8) is the same as that for (16).*

Proof: It needs to show for any $\pi^* \in \Pi^*(\beta_0)$ that π^* is an optimal solution to (16), and vice versa. By Theorem 1 and (8) there is

$$\begin{cases} 0 = \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r^\pi(\zeta) \\ \pi^* \in \operatorname{argmin}_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta^* r^*(\zeta) - r^\pi(\zeta) \end{cases}$$

$$\Leftrightarrow \begin{cases} 0 = \max_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r^{\pi^*}(\zeta) \\ \forall \pi \in \Pi\ 0 \leq \max_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r^\pi(\zeta) \end{cases}$$

$$\Leftrightarrow \begin{cases} \exists \zeta \in \mathcal{U}\ 0 = \beta_0 r^*(\zeta) - r^{\pi^*}(\zeta) \\ \forall \zeta \in \mathcal{U}\ 0 \geq \beta_0 r^*(\zeta) - r^{\pi^*}(\zeta) \\ \forall \pi \in \Pi\ \exists \zeta \in \mathcal{U}\ 0 \leq \beta_0 r^*(\zeta) - r^\pi(\zeta) \end{cases}$$

$$\Leftrightarrow \begin{cases} \exists \zeta \in \mathcal{U} : \beta_0 = r^{\pi^*}(\zeta)/r^*(\zeta) \\ \forall \zeta \in \mathcal{U} : \beta_0 \leq r^{\pi^*}(\zeta)/r^*(\zeta) \\ \forall \pi \in \Pi\ \exists \zeta \in \mathcal{U}\ \beta_0 \geq r^\pi(\zeta)/r^*(\zeta) \end{cases}$$

$$\Leftrightarrow \begin{cases} \beta_0 = \min_{\zeta \in \mathcal{U}} r^{\pi^*}(\zeta)/r^*(\zeta) \\ \forall \pi \in \Pi\ \beta_0 \geq \min_{\zeta \in \mathcal{U}} r^\pi(\zeta)/r^*(\zeta) \end{cases}$$

$$\Leftrightarrow \begin{cases} \beta^* = \max_{\pi \in \Pi} \min_{\zeta \in \mathcal{U}} r^\pi(\zeta)/r^*(\zeta) \\ \pi^* \in \operatorname{argmax}_{\pi \in \Pi} \min_{\zeta \in \mathcal{U}} r^\pi(\zeta)/r^*(\zeta) \end{cases}$$

As the reasoning can go in both directions, the theorem is established. ∎

Based on the result of Lemma 4, the next lemma gives the condition for the existence of a unique competitive ratio.

Lemma 5. *If $D(0) < 0$ then there is $r^*(\zeta) > 0$ for all $\zeta \in \mathcal{U}$, and there is a unique $\beta_0 \in (0, 1]$ such that $D(\beta_0) = 0$.*

Proof: Note that at $\beta = 0$ it becomes equivalent to the maximin reward criterion:

$$D(0) = \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} -r^\pi(\zeta) = -\max_{\pi \in \Pi} \min_{\zeta \in \mathcal{U}} r^\pi(\zeta).$$

21

Suppose there is a $\dot{\zeta}$ such that $r^*(\dot{\zeta}) \leq 0$, then there is

$$-D(0) = \max_{\pi \in \Pi} \min_{\zeta \in \mathcal{U}} r^\pi(\zeta) \leq \max_{\pi \in \Pi} r^\pi(\dot{\zeta}) = r^*(\dot{\zeta}) \leq 0.$$

Therefore $D(0) \geq 0$, a contradiction! Thus there is $r^*(\zeta) > 0$ for all $\zeta \in \mathcal{U}$, so $D(\beta)$ strictly increases in β. Note that at $\beta = 1$ it is the minimax regret criterion, thus $D(1) \geq 0 > D(0)$, and the conclusion follows by the monotonicity and continuity of $D(\beta)$ inferred from Lemma 2 and Theorem 2. ∎

Variants of Lemma 4 and 5 are known for fractional combinatorial optimization, see e.g. Megiddo (1978) for numerical algorithms based on them. They are also useful to analytically solve for competitive ratios, as will be illustrated with the one-way trading problem later.

4. Tractability.

This section deals with the tractability of two-stage linear RO problems with an ARM criterion by converting them into the following problem of two-stage linear RO with fixed recourse (TSLRO/FR):

$$\max_{x \in \mathcal{X}, y(\cdot)} \inf_{\zeta \in \mathcal{U}} \quad (C\zeta + c)^T x + d^T y(\zeta) + f^T \zeta,$$
$$\text{s.t.} \quad Ax + By(\zeta) \leq \Psi\zeta + \psi, \forall \zeta \in \mathcal{U},$$

where $x \in \mathbb{R}^{n_x}$ is the first-stage action before the revelation of the uncertain parameters $\zeta \in \mathbb{R}^{n_\zeta}$, while $y : \mathbb{R}^{n_\zeta} \to \mathbb{R}^{n_y}$ is a strategy for the second-stage action implemented after ζ has been revealed. The constants are $\Psi \in \mathbb{R}^{m \times n_\zeta}$, $C \in \mathbb{R}^{n_x \times n_\zeta}, c \in \mathbb{R}^{n_x}, d \in \mathbb{R}^{n_y}, f \in \mathbb{R}^{n_\zeta}, A \in \mathbb{R}^{m \times n_x}$ and $B \in \mathbb{R}^{m \times n_y}$. Both $\mathcal{X}$ and $\mathcal{U}$ are nonempty polyhedra: $\mathcal{X} := \{x \in \mathbb{R}^{n_x} | Wx \leq v\}$ with $W \in \mathbb{R}^{r \times n_x}$

22

and $v \in \mathbb{R}^r$, and $\mathcal{U} := \{\zeta \in \mathbb{R}^{n_\zeta} | P\zeta \leq q\}$ with $P \in \mathbb{R}^{s \times n_\zeta}$ and $q \in \mathbb{R}^s$. The more common notation of $\min_{\zeta \in \mathcal{U}}$ may be used if $\mathcal{U}$ is bounded. The two stages of decisions can be separated to have

$$\max_{x \in \mathcal{X}} \inf_{\zeta \in \mathcal{U}} h(x, \zeta)$$

where $h(x, \zeta)$ is the second-stage optimal objective found in the linear *recourse problem* defined as

$$
\begin{aligned}
h(x, \zeta) \quad := \quad &\sup_{y} (C\zeta + c)^T x + d^T y + f^T \zeta \qquad\qquad (17) \\
s.t. \quad : \quad &Ax + By \leq \Psi(x)\zeta + \psi.
\end{aligned}
$$

The fixed recourse property refers to d and B being unaffected by ζ, which is essential for approximate schemes using linear decision rules in the form of an affine policy $y(\zeta) := \Upsilon\zeta + y$, where $\Upsilon \in \mathbb{R}^{n_y \times n_\zeta}$ and $y \in \mathbb{R}^{n_y}$.

The seminal work of Ben-Tal et al. (2004) establishes that the TSLRO/FR problem is NP-hard in general, since the so-called "adversarial problem" of $\inf_{\zeta \in \mathcal{U}} h(x, \zeta)$ minimizes a piecewise linear concave function over an arbitrary polyhedron and is NP-hard in itself. A tractable approximation of the TSLRO/FR problem initially proposed in Ben-Tal et al. (2004)) employs linear decision rules for the second-stage strategy $y(\cdot)$, which can be reformulated into an LP model by exploiting the principles of duality theory. In the last decade, a number of theoretical and empirical studies have shown that linear decision rules can provide high-quality solutions to TSLRO/FR problems. Furthermore, Bertsimas et al. (2010), Ardestani-Jaafari and Delage (2016), and Poursoltani and Delage (2021) establish conditions under which this approach is exact. Methods to identify exact solutions are also developed, among which is the Column-and-Constraint Generation method

proposed by Zeng and Zhao (2013). The reader is referred to Delage and Iancu (2015) and Yanıkoğlu et al. (2019) for a rich set of additional methods to solve TSLRO/FR problems efficiently.

To leverage these methods for TSLRO/FR problems, it is highly desirable to convert two-stage linear RO with ARM (LROARM) problems into equivalent TSLRO/FR problems. Given the optimal second-stage profit $h(x, \zeta)$, the LROARM problem takes the form

$$\min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) - h(x, \zeta), \tag{18}$$

which is well-defined when the best profit achievable in hindsight never reaches infinity, that is, $\sup_{x' \in \mathcal{X}} h(x', \zeta) < \infty, \forall \zeta \in \mathcal{U}$. Equivalent reformulation into TSLRO/FR problems will be considered next for the LROARM problem with either right-hand side or objective uncertainty. The following assumptions are frequently used later on.

Assumption 1 (Existence). *The sets $\mathcal{X}$ and $\mathcal{U}$ are nonempty polyhedra, and there exists a triplet (x, ζ, y) such that $x \in \mathcal{X}, \zeta \in \mathcal{U}$, and $Ax + By \leq \Psi(x)\zeta + \psi$.*

Assumption 2 (Relatively Complete Recourse). *For all $x \in \mathcal{X}$ and for all $\zeta \in \mathcal{U}$, there always exists a recourse action y to satisfy all the constraints,*

$$\forall x \in \mathcal{X}, \forall \zeta \in \mathcal{U}, \exists y \in \mathbb{R}^{n_y}, Ax + By \leq \Psi(x)\zeta + \psi. \tag{19}$$

Assumption 3 (Finite Case). *For all $x \in \mathcal{X}$ there exists a $\zeta \in \mathcal{U}$ such that $h(x, \zeta)$ is bounded from above. Equivalently, there exists a function $\bar{\zeta} : \mathcal{X} \to \mathcal{U}$ such that $\forall x \in \mathcal{X}, h(x, \bar{\zeta}(x)) < \infty$.*

Assumption 4 (Finite Worst-Case). *There is a lower bound on the worst-case profit achievable:* $\forall x \in \mathcal{X}, \inf_{\zeta \in \mathcal{U}} h(x, \zeta) > -\infty.$

Assumption 5 (Finite Best-Case). *There is an upper bound on the best-case profit achievable:* $\sup_{x \in \mathcal{X}, \zeta \in \mathcal{U}} h(x, \zeta) < \infty.$

Assumption 5 is a natural condition to impose on LROARM problems and implies Assumption 3.

4.1. Right-Hand Side Uncertainty

This subsection deals with the case with the uncertainty limited to the right-hand side, where the profit function $h(x, \zeta)$ takes the following form,

$$h(x, \zeta) \quad := \quad \sup_{y} c^T x + d^T y \tag{20}$$

$$s.t. \quad : \quad Ax + By \le \Psi \zeta + \psi. \tag{21}$$

Let $Y(x, \zeta) := \{y \in \mathbb{R}^{n_y} | Ax + By \le \Psi\zeta + \psi\}$ and consider the two consecutive sup operators in (18) together with (20) to have

$$\sup_{\zeta \in \mathcal{U}} \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) \quad = \quad \sup_{\zeta \in \mathcal{U}} \sup_{x' \in \mathcal{X}} \sup_{y' \in Y(x', \zeta)} \beta(c^T x' + d^T y')$$

$$= \quad \sup_{\zeta' \in \mathcal{U}'} \beta(c^T x' + d^T y')$$

where $\zeta' = [\zeta^T \ x'^T \ y'^T]^T$ is called a lifting of ζ as a result of merging the three consecutive sup operators, and $\mathcal{U}' = \{[\zeta^T \ x'^T \ y'^T]^T : \zeta \in \mathcal{U}, x' \in \mathcal{X}, y' \in Y(x', \zeta)\}$ is the lifted uncertainty set.

Theorem 5. *Given Assumption 1, the LROARM problem with right-hand side uncertainty is equivalent to the following TSLRO/FR problem:*

$$- \max_{x \in \mathcal{X}, y(\cdot)} \inf_{\zeta' \in \mathcal{U}'} \quad c^T x + d^T y(\zeta') + \beta f'^T \zeta', \tag{22}$$

$$s.t. \quad Ax + By(\zeta') \le \Psi'\zeta' + \psi, \forall \zeta' \in \mathcal{U}', \tag{23}$$

25

where $\zeta' = [\zeta^T \ x'^T \ y'^T]^T \in \mathbb{R}^{n_\zeta + n_x + n_y}$, $y : \mathbb{R}^{n_\zeta + n_x + n_y} \to \mathbb{R}^{n_y}$, $f' := [0^T \ -$

$c^T \ -d^T]^T$, $\Psi' := [\Psi \ 0 \ 0]$, *and the lifted uncertainty set* $\mathcal{U}'$ *can be defined by*

$$\mathcal{U}' := \{\zeta' \in \mathbb{R}^{n_\zeta + n_x + n_y} | P'\zeta' \leq q'\}, \ with \tag{24}$$

$$P' = \begin{bmatrix} P & 0 & 0 \\ 0 & W & 0 \\ -\Psi & A & B \end{bmatrix} \ and \ q' = \begin{bmatrix} q \\ v \\ \psi \end{bmatrix}.$$

Furthermore, Assumption 1 carries through the reformulation naturally, and so does Assumption 2. Assumption 3 carries through if Assumptions 2 also holds for the LROARM problem. And finally, Assumption 4 carries through if Assumptions 5 also holds for the LROARM problem.

Proof: Start from the definition of LROARM with right-hand side uncertainty and proceed with the following simple steps:

$$\begin{aligned} \text{LROARM} \quad &\equiv \quad \min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \left\{ \sup_{x' \in \mathcal{X}, y' \in Y(x', \zeta)} \beta(c^T x' + d^T y') - \sup_{y \in Y(x, \zeta)} c^T x + d^T y \right\}, \\ &\equiv \quad \min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}, x' \in \mathcal{X}, y' \in Y(x', \zeta)} \inf_{y \in Y(x, \zeta)} \beta(c^T x' + d^T y') - (c^T x + d^T y), \\ &\equiv \quad -\max_{x \in \mathcal{X}} \inf_{\zeta \in \mathcal{U}, x' \in \mathcal{X}, y' \in Y(x', \zeta)} \sup_{y \in Y(x, \zeta)} c^T x + d^T y - \beta(c^T x' + d^T y'), \\ &\equiv \quad -\max_{x \in \mathcal{X}, y(\cdot)} \inf_{\zeta' \in \mathcal{U}'} c^T x + d^T y(\zeta') + \beta f'^T \zeta', \\ &\qquad \text{s.t. } Ax + By(\zeta') \leq \Psi'\zeta' + \psi, \forall \zeta' \in \mathcal{U}', \end{aligned}$$

where $Y(x, \zeta) := \{y \in \mathbb{R}^{n_y} | Ax + By \leq \Psi\zeta + \psi\}$, and where the minimization and maximization operations are simply regrouped, and then the sign of the objective is flipped.

It remains to verify the conditions under which the assumptions are satisfied by this new TSLRO/FR problem. First, when Assumption 1 holds

for the LROARM problem, there must exist a triplet $(\bar{x}, \bar{\zeta}, \bar{y})$ satisfying $\bar{x} \in \mathcal{X}, \bar{\zeta} \in \mathcal{U}$, and $bary \in Y(\bar{x}, \bar{\zeta})$. It is clear that $\zeta' := [\bar{\zeta}^T \ \bar{x}^T \ \bar{y}^T]^T \in \mathcal{U}'$ so that the triplet $(\bar{x}, \zeta', \bar{y})$ satisfies Assumption 1 for the reformulation. There-fore, Assumption 1 carries through naturally. Second, since the feasible set for the recourse problem is the same in LROARM and its TSLRO/FR re-formulation, Assumption 2 carries also carries through unscathed. Third, one can show that Assumption 3 also carries through if Assumption 2 holds. Simply let $\bar{\zeta} : \mathcal{X} \to \mathcal{U}$ map an x to a ζ that must exist according to Assump-tion 3 for the LROARM problem, and let (x', y') be a feasible first-stage and recourse action, which exists based on Assumption 2. Then the mapping $\zeta'(x) := [\bar{\zeta}^T(x) \ x'^T \ y'^T]^T$ provides that special ζ' to satisfy Assumption 3 for the TSLRO/FR problem. Finally, Assumption 4 carries through as long as Assumption 5 also holds for the LROARM problem. With Assumption 4 and 5 satisfied by the LROARM problem, let the recourse problem that appears in the TSLRO/FR reformulation as $h'(x, \zeta')$,

$$
\begin{aligned}
\inf_{\zeta' \in \mathcal{U}'} h'(x, \zeta') &= \inf_{\zeta \in \mathcal{U}, x' \in \mathcal{X}, y' \in Y(x', \zeta)} \ \sup_{y \in Y(x, \zeta)} c^T x + d^T y - \beta(c^T x' + d^T y'), \\
&\geq \inf_{\zeta \in \mathcal{U}} h(x, \zeta) - \sup_{x' \in \mathcal{X}, y' \in Y(x', \zeta)} \beta(c^T x' + d^T y'), \\
&\geq \inf_{\zeta \in \mathcal{U}} h(x, \zeta) - \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) > -\infty.
\end{aligned}
$$

∎

4.2. Objective Uncertainty

This subsection deals with the case when uncertainty is limited to the ob-jective function. To be precise, the profit function $h(x, \zeta)$ takes the following

form,

$$h(x, \zeta) \quad := \quad \sup_{y} c^T x + d^T(\zeta) y \tag{25}$$

$$s.t. \quad : \quad Ax + By \leq \psi. \tag{26}$$

Note that this concise form help simplify exposition without losing the more general form where c is uncertain, which can be accommodated by lifting the space of second-stage decisions:

$$h(x, \zeta) \quad := \quad \sup_{y_x, y_y} c^T(\zeta) y_x + d^T(\zeta) y_y$$

$$s.t. \quad : \quad y_x = x$$

$$Ax + By_y \leq \psi.$$

Theorem 6. *Given Assumptions 1 and 2, the LROARM problem with objective uncertainty is equivalent to the following TSLRO/FR problem:*

$$- \max_{x \in \mathcal{X}, y'(\zeta')} \inf_{\zeta' \in \mathcal{U}'} \quad (C\zeta' + c)^T x + \beta d'^T y'(\zeta') + f'^T \zeta', \tag{27}$$

$$s.t. \quad A'x + B'y'(\zeta') \leq \Psi'\zeta' + \psi',$$

where $y' : \mathbb{R}^{n_\varsigma + m} \to \mathbb{R}^{m+r}$, whereas $\mathcal{U}'$ is the new uncertainty set defined as:

$$\mathcal{U}' := \{\zeta' \in \mathbb{R}^{n_\varsigma + m} | P'\zeta' \leq q'\}, \quad with$$

$$P' = \begin{bmatrix} P & 0 \\ -D & B^T \\ D & -B^T \end{bmatrix} \quad and \quad q' = \begin{bmatrix} q \\ d \\ -d \end{bmatrix}$$

and where the matrices are constructed as follows,

$$C' := \begin{bmatrix} 0 & -A^T \end{bmatrix}, d' := \begin{bmatrix} -\psi \\ -v \end{bmatrix}, f' := \begin{bmatrix} 0 \\ \psi \end{bmatrix}, A' := \begin{bmatrix} 0 \end{bmatrix},$$

$$B' := \begin{bmatrix} A^T & W^T \\ -A^T & -W^T \\ B^T & 0 \\ -B^T & 0 \\ -I & 0 \\ 0 & -I \end{bmatrix}, \Psi' := \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ D & 0 \\ -D & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \text{ and } \psi' := \begin{bmatrix} c \\ -c \\ d \\ -d \\ 0 \\ 0 \end{bmatrix}.$$

Furthermore, the TSLRO/FR reformulation satisfies Assumptions 1 and 2 when the LROARM also satisfies Assumptions 3 and 5, whereas the LROARM needs to additionally satisfy Assumption 4 for the TSLRO/FR reformulation to satisfy Assumptions 3 and 4.

Proof: Consider the ex post problem in the LROARM problem with objective uncertainty:

$$\sup_{x' \in \mathcal{X}} h(x', \zeta) = \sup_{x', y'} c^T x' + d^T(\zeta) y', \tag{28}$$

$$\text{s.t.} \quad Ax' + By' \leq \psi, \tag{29}$$

$$Wx' \leq v, \tag{30}$$

to which Assumption 2 guarantees a feasible solution (x', y'). Therefore, strong duality holds and the dual form is derived as

$$\sup_{x' \in \mathcal{X}} h(x', \zeta) = \inf_{\lambda \geq 0, \gamma \geq 0} \psi^T \lambda + v^T \gamma, \tag{31}$$

$$\text{s.t.} \quad A^T \lambda + W^T \gamma = c, \tag{32}$$

$$B^T \lambda = d(\zeta), \tag{33}$$

where the dual variables $\lambda \in \mathbb{R}^m$ and $\gamma \in \mathbb{R}^r$ are associated with constraints (29) and (30). Similarly, Assumption 2 also provides strong duality for the recourse problem (17). Thus it is possible to substitute both $h(x, \zeta)$ and $\sup_{x' \in \mathcal{X}} h(x', \zeta)$ in the LROARM problem by their respective dual form:

$$
\begin{aligned}
& \min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) - h(x, \zeta), \\
\equiv \; & \min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \left\{ \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) - \inf_{\rho \in \Phi(\zeta)} c^T x + (\psi - Ax)^T \rho \right\}, \\
\equiv \; & \min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}, \rho \in \Phi(\zeta)} \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) - (c^T x + (\psi - Ax)^T \rho), \\
\equiv \; & \min_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}, \rho \in \Phi(\zeta)} \inf_{(\lambda, \gamma) \in \Lambda(\zeta)} \beta(\psi^T \lambda + v^T \gamma) - (c^T x + (\psi - Ax)^T \rho), \\
\equiv \; & -\max_{x \in \mathcal{X}} \inf_{\zeta \in \mathcal{U}, \rho \in \Phi(\zeta)} \sup_{(\lambda, \gamma) \in \Lambda(\zeta)} c^T x + (\psi - Ax)^T \rho - \beta(\psi^T \lambda + v^T \gamma), \quad (34)
\end{aligned}
$$

where $\Lambda(\zeta) := \{(\lambda, \gamma) \in \mathbb{R}^m \times \mathbb{R}^r | (32), (33), \lambda, \gamma \geq 0\}$ and $\Phi(\zeta) := \{\rho \in \mathbb{R}^m | B^T \rho = d(\zeta), \rho \geq 0\}$. By having $\zeta' := [\zeta^T \; \rho^T]^T$ and $y' := [\lambda^T \; \gamma^T]^T$, problem (34) can be rewritten in the reformulation (27).

It remains to verify the conditions on LROARM under which the assumptions are satisfied by this new TSLRO/FR problem. First consider that LROARM satisfies Assumptions 1, 2, 3, and 5. Based on Assumption 3, for all $x \in \mathcal{X}$ there is $h(x, \bar{\zeta}(x)) < \infty$. This implies by LP duality that there must be a feasible $\bar{\rho} \in \Phi(\bar{\zeta}(x))$. Moreover, Assumption 5 implies that $\sup_{x' \in \mathcal{X}} h(x', \bar{\zeta}(x)) < \infty$ hence once again LP duality ensures that there exists a pair $(\bar{\lambda}, \bar{\gamma}) \in \Lambda(\bar{\zeta}(x))$. The TSLRO/FR reformulation therefore satisfies Assumption 1 using $(x, \bar{\zeta}(x), \bar{\rho}, \bar{\lambda}, \bar{\gamma})$. Next, the fact that the TSLRO/FR reformulation satisfies Assumption 2 follows similarly from imposing Assumption 5 on LROARM because the existence of a pair $\bar{\lambda}, \bar{\gamma}) \in \Lambda(\zeta)$ holds for all $\zeta \in \mathcal{U}$.

Now consider that LROARM satisfies in addition Assumption 4. Assumption 3 and 4 on LROARM implies that there exists a $\bar{\zeta}(x) \in \mathcal{U}$ such that, for all $x \in \mathcal{U}, \infty > h(x, \bar{\zeta}(x)) \geq \inf_{\zeta \in \mathcal{U}} > -\infty$. Therefore,

$$
\begin{aligned}
\inf_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) - h(x, \zeta) &\geq \inf_{x \in \mathcal{X}} \sup_{x' \in \mathcal{X}} \beta h(x', \bar{\zeta}(x)) - h(x, \bar{\zeta}(x)) \\
&\geq \beta h(x, \bar{\zeta}(x)) - h(x, \bar{\zeta}(x)) = (\beta - 1)h(x, \bar{\zeta}(x)) > -\infty.
\end{aligned}
$$

The LROARM problem is therefore bounded below hence the TSLRO/FR reformulation is bounded above, which demonstrates that the latter satisfies Assumption 3. Similarly, there is for all $x \in \mathcal{X}$:

$$
\sup_{\zeta \in \mathcal{U}} \left\{ \sup_{x' \in \mathcal{X}} h(x', \zeta) - h(x, \zeta) \right\} \leq \sup_{\zeta \in \mathcal{U}} \sup_{x' \in \mathcal{X}} h(x', \zeta) - \inf_{\zeta \in \mathcal{U}} h(x, \zeta) < \infty,
$$

where the first term is bounded above according to Assumption 5 and the second term bounded below according to Assumption 4. We can thus conclude that for all $x \in \mathcal{X}$, the worst-case regret is bounded above; and that for all $x \in \mathcal{X}$, the worst-case profit in the TSLRO/FR reformulation is bounded below, that is, Assumption 4 is satisfied by the TSLRO/FR reformulation. ■

With LROARM problems reformulated to equivalent TSLRO/FR problems, the approximate and exact methods developed for TSLRO/FR problems can be readily leveraged to solve LROARM problems, as well as the theoretical results for linear decision rules to be exact.

5. One-way Trading.

The one-way trading problem is studied here with the ARM criterion for multiple demonstrations, for which it is an ideal choice, especially that it is

already studied under RO with both competitive ratio (El-Yaniv et al. 2001) and absolute regret (Wang et al. 2016). First, the ARM criterion is amenable to analysis by producing closed-form solutions, which yields the result of Wang et al. (2016) as a special case at $\beta = 1$. Second, the newly proposed approach to competitive ratio analysis is applied, without depending on acute intuition and insights as in El-Yaniv et al. (2001), which drastically reduces the difficulty of analysis. Third, the solution indeed becomes more aggressive as β increases, and the effects of β on the performance can be shown via numerical simulations. These are helpful to demonstrate the properties and potential of the ARM criterion.

5.1. Problem Formulation.

Consider selling a certain amount of fully divisible goods (like gasoline or steel) in a finite time horizon, while the price fluctuates in the range of $[m, M]$. For comparable results, the tradition of dividing time into T discrete periods is followed. A fixed price $p_t \in [m, M]$ is revealed in each period $t = 1, \cdots, T$. The trader is a price-taker and must decide in each period on the amount x_t to sell at p_t without knowing the future prices. The goal is to maximize the total sales revenue in the end.

It is helpful to connect to the notations in section 2. A scenario ζ simply corresponds to the prices $p = (p_1, \cdots, p_T)$ revealed over time, with $\zeta_t = p_t$. There is $\mathcal{U} = [m, M]^T$ and $\mathcal{U}_t(\zeta_{1:t}) = [m, M]$, as prices are independent of each other. Without loss of generality, the total amount of goods to sell is one unit, and the action is $x = (x_1, \cdots, x_T)$ with $X = \{x : \sum_{t=1}^{T} x_t = 1, x \geq 0\}$. For $t < T$ there is $X_t(h_t) = [0, q_t]$ where $q_t = 1 - \sum_{s=1}^{t-1} x_s$ is the remaining amount to sell given h_t, but in the last period everything

32

must go, so $X_T(h_T) = [q_T, q_T]$. The reward is accumulated over time, so let $r_t = \sum_{s=1}^{t-1} p_s x_s$ be the rewards accumulated in h_t, the reward in the end is $r(x, p) = r_{T+1}$. Let $\hat{p}_t = \max\{p_s : s = 1, \cdots, t-1\}$ denote the highest price seen in h_t, and $r^*(p) = \max\{r(x, p) : \sum_{t=1}^{T} x_t = 1\} = \hat{p}_{T+1}$ the ex post optimal. At the end of the last stage (2) becomes

$$D_T(h_{T+1}; \beta) = \beta \hat{p}_{T+1} - r_{T+1}. \tag{35}$$

In this multistage problem it is natural to have periods coincide with stages, in which the uncertain price is first revealed, then an action is taken. This calls for a different formulation from the standard formulation in (4):

$$D_{t-1}(h_t; \beta) = \max_{p_t \in [m,M]} \min_{x_t \in X_t(h_t)} D_t(h_{t+1}; \beta), \tag{36}$$

but the difference is only superficial: all the results in section 3 remain valid.

5.2. Analytic Solution.

The analysis starts from the last period T and works backwards. In the last period clearly there is $x_T = q_T$, and (36) becomes

$$D_{T-1}(h_T; \beta) = \max_{p_T \in [m,M]} \beta \max(\hat{p}_T, p_T) - (r_T + p_T q_T),$$

which is convex in p_T, and the maximizer is either $p_T = m$ or $p_T = M$. Define auxiliary functions that map a quantity $q \in [0, 1]$ to a price in $[m, M]$,

$$P_j(q) = (M - m)\left(1 - \frac{q}{\beta j}\right)^{+j} + m, j = 1, 2, \cdots,$$

where $y^{+j} = \max^j(0, y)$ denote the positive part of y raised to the jth power. Let $P_j^-(y) = q$ be the inverse of $y = P_j(q)$ for $q \in [0, \beta j]$.

$$
\begin{aligned}
D_{T-1}(h_T; \beta) &= \max(\beta \hat{p}_T - R_T, \beta M - (r_T + Mq_T)) \\
&= \max(\beta \hat{p}_T, \beta M - (M - m)q_T) - R_T \\
&= \beta \max(\hat{p}_T, P_1(q_T)) - R_T,
\end{aligned}
$$

where $R_t = r_t + mq_t$ for $t = 1, \cdots, T$ is the lower bound on r_{T+1} given h_t. Note that the trivial case of $\beta = 0$ is not considered. Continue on with (36) for $t = T - 1, \cdots, 1$, the result is obtained and presented as follows.

Theorem 7. *The minimal worst-case regret for the one-way trading problem in period t given history h_t for $t = 1, 2, \cdots, T$ is*

$$
D_{t-1}(h_t; \beta) = \beta \max(\hat{p}_t, P_{1+T-t}(q_t)) - R_t, \tag{37}
$$

and the optimal trading policy is $\pi_t^*(h_t, p_t) = q_t - q_{t+1}^*$, where $q_{T+1}^* = 0$ and

$$
q_{t+1}^* = \min(q_t, P_{T-t}^-(\hat{p}_{t+1})), \quad t = 1, \cdots, T - 1. \tag{38}
$$

Proof: By backward induction. It is already verified for period $t = T$, which completes the initial step. For the induction step, assume that (37) holds in period $t + 1 (t < T)$, that is,

$$
D_t(h_{t+1}; \beta) = \beta \max(\hat{p}_{t+1}, P_{T-t}(q_{t+1})) - R_{t+1},
$$

and show that it also holds in period t. For the minimization nested in (36), let

$$
\begin{aligned}
\bar{D}_t(h_t, p_t; \beta) &= \min_{x_t \in X_t(h_t)} D_t(h_{t+1}; \beta) \\
&= \min_{q_{t+1} \in [0, q_t]} \beta \max(\hat{p}_{t+1}, P_n(q_{t+1})) - R_{t+1}, \tag{39}
\end{aligned}
$$

34

with $n = T-t$, $q_{t+1} = q_t - x_t$, and $R_{t+1} = r_{t+1} + mq_{t+1}$. To find $\partial D_t(h_{t+1}; \beta)/\partial q_{t+1}$, first note that

$$P_n'(q) = -\frac{M-m}{\beta}\left(1 - \frac{q}{\beta n}\right)^{+(n-1)} \leq 0.$$

Then by the monotonicity of $P_n(q)$, there is $P_n(q_{t+1}) \geq \hat{p}_{t+1}$ if $q_{t+1} \leq P_n^-(\hat{p}_{t+1})$, and similarly $q_{t+1} > P_n^-(\hat{p}_{t+1})$ ensures $P_n(q_{t+1}) \leq \hat{p}_{t+1}$. Thus there is

$$D_t(h_{t+1}; \beta) = \begin{cases} \beta P_n(q_{t+1}) - R_{t+1} & q_{t+1} \leq P_n^-(\hat{p}_{t+1}) \\ \beta \hat{p}_{t+1} - R_{t+1} & q_{t+1} > P_n^-(\hat{p}_{t+1}) \end{cases} \tag{40}$$

$$\frac{\partial D_t(h_{t+1}; \beta)}{\partial q_{t+1}} = \begin{cases} p_t - m + \beta P_n'(q_{t+1}) & q_{t+1} < P_n^-(\hat{p}_{t+1}) \\ p_t - m & q_{t+1} > P_n^-(\hat{p}_{t+1}) \end{cases} \tag{41}$$

Note that in the first branch with $q_{t+1} < P_n^-(\hat{p}_{t+1})$, there is $p_t \leq \hat{p}_{t+1} < P_n(q_{t+1}) \leq -\beta P_n'(q_{t+1}) + m$, so $p_t - m + \beta P_n'(q_{t+1}) < 0$. And in the second branch with $q_{t+1} > P_n^-(\hat{p}_{t+1})$, there is $p_t - m \geq 0$. Therefore an optimal solution to (39) is (38), which from (40) gives

$$\bar{D}_t(h_t, p_t; \beta) = \beta P_n(q_{t+1}^*) - (r_{t+1} + mq_{t+1}^*). \tag{42}$$

Let $\bar{p}_t = \max(\hat{p}_t, P_n(q_t)) \in [m, M]$, and from (36) there is

$$\begin{aligned} D_{t-1}(h_t; \beta) &= \max_{p_t \in [m,M]} \bar{D}_t(h_t, p_t; \beta) \\ &= \max\left(\begin{array}{c} \max_{p_t \in [m,\bar{p}_t]} \bar{D}_t(h_t, p_t; \beta) \\ \max_{p_t \in [\bar{p}_t,M]} \bar{D}_t(h_t, p_t; \beta) \end{array}\right) \end{aligned} \tag{43}$$

For the branch with $p_t \in [m, \bar{p}_t]$ in (43), consider two cases: (i) $\bar{p}_t = \hat{p}_t \geq P_n(q_t)$ and (ii) $\bar{p}_t = P_n(q_t) > \hat{p}_t$. In case (i) there is $\hat{p}_{t+1} = \max(\hat{p}_t, p_t) = \hat{p}_t \geq P_n(q_t)$, therefore $P_n^-(\hat{p}_{t+1}) \leq q_t$ and (38) simplifies to $q_{t+1}^* = P_n^-(\hat{p}_{t+1})$,

35

thus $P_n(q^*_{t+1}) = \hat{p}_{t+1} = \bar{p}_t$. In case (ii) there is $\hat{p}_{t+1} \leq P_n(q_t)$, therefore $P_n^-(\hat{p}_{t+1}) \geq q_t$ and (38) simplifies to $q^*_{t+1} = q_t$, thus $P_n(q^*_{t+1}) = P_n(q_t) = \bar{p}_t$. As in both cases there is $P_n(q^*_{t+1}) = \bar{p}_t$, from (42) there is $\bar{D}_t(h_t, p_t; \beta) = \beta\bar{p}_t - (r_{t+1} + mq^*_{t+1}) = \beta\bar{p}_t - r_t - p_t x^*_t - mq^*_{t+1}$, which is linear in p_t with a slope of $-x^*_t \leq 0$ as $x^*_t = q_t - q^*_{t+1} \geq 0$. Thus $p^*_t = m$ is a maximizer, which gives

$$\max_{p_t \in [m, \bar{p}_t]} \bar{D}_t(h_t, p_t; \beta) = \beta\bar{p}_t - r_t - mq_t = \beta\bar{p}_t - R_t.$$

For the branch in (43) with $p_t \in [\bar{p}_t, M]$, as $p_t \geq \bar{p}_t \geq \hat{p}_t$, there is $\hat{p}_{t+1} = p_t \geq \bar{p}_t \geq P_n(q_t)$, thus $P_n^-(\hat{p}_{t+1}) \leq q_t$ and (38) simplifies to $q^*_{t+1} = P_n^-(\hat{p}_{t+1})$. Therefore $P_n(q^*_{t+1}) = \hat{p}_{t+1} = p_t$, and (42) simplifies to $\bar{D}_t(h_t, p_t; \beta) = \beta p_t - r_t - p_t x^*_t - mq^*_{t+1} = \beta p_t - r_t - p_t(q_t - q^*_{t+1}) - mq^*_{t+1} = (\beta - q_t + q^*_{t+1})p_t - mq^*_{t+1} - r_t = (\beta - q_t + q^*_{t+1})P_n(q^*_{t+1}) - mq^*_{t+1} - r_t$. Now consider this function

$$d(z) = (\beta - q_t + z)P_n(z) - mz - r_t, z \in [0, 1].$$

The derivative is $d'(z) = (\beta - q_t + z)P'_n(z) + P_n(z) - m$. Note that $P_n(z) - m = -(\beta - z/n)P'_n(z)$, thus $d'(z) = (\beta - q_t + z)P'_n(z) - (\beta - z/n)P'_n(z) = (z + z/n - q_t)P'_n(z)$. As $P'_n(z) \leq 0$, there is $d'(z) \geq 0$ when $z + z/n - q_t \leq 0$, and $d'(z) \leq 0$ when $z + z/n - q_t \geq 0$, hence $z^* = nq_t/(n+1)$ is a maximizer of $d(z)$, which gives

$$d(z^*) = \beta P_{n+1}(q_t) - R_t,$$
$$P_n(z^*) \geq P_{n+1}(q_t).$$

Clearly, there is $\bar{D}_t(h_t, p_t; \beta) = d(P_n^-(p_t))$ for $p_t \in [\bar{p}_t, M]$, consider two cases. Case (i) $P_n(z^*) \geq \bar{p}_t$. As $P_n^-(M) = 0 \leq z^* \leq P_n^-(\bar{p}_t)$, there is $\max_{p_t \in [\bar{p}_t, M]} \bar{D}_t(h_t, p_t; \beta) = d(z^*)$. Thus, according to (43) there is

$$D_{t-1}(h_t; \beta) = \max(\beta\bar{p}_t - R_t, d(z^*)).$$

36

Case (ii) $P_n(z^*) < \bar{p}_t$. As $q_t \geq z^* \geq P_n^-(\bar{p}_t)$, there is

$$\max_{p_t \in [\bar{p}_t, M]} \bar{D}_t(h_t, p_t; \beta) = \max_{p_t \in [\bar{p}_t, M]} d(P_n^-(p_t))$$

$$= \max_{z \in [0, P_n^-(\bar{p}_t)]} d(z)$$

$$\leq \max_{z \in [0, q_t]} d(z) = d(z^*).$$

As $P_n(z^*) \geq P_{n+1}(q_t)$, there is $\bar{p}_t \geq P_{n+1}(q_t)$. Therefore $d(z^*) = \beta P_{n+1}(q_t) - R_t \leq \beta \bar{p}_t - R_t$, and according to (43) there is,

$$D_{t-1}(h_t; \beta) = \beta \bar{p}_t - R_t = \max(\beta \bar{p}_t - R_t, d(z^*)).$$

So in both cases there is $D_{t-1}(h_t; \beta) = \max(\beta \bar{p}_t - R_t, d(z^*))$. Note that $\bar{p}_t = \max(\hat{p}_t, P_n(q_t))$, and $P_n(q_t) \leq P_{n+1}(q_t)$, thus

$$\begin{aligned}
D_{t-1}(h_t; \beta) &= \max(\beta \bar{p}_t - R_t, d(z^*)) \\
&= \max(\beta \bar{p}_t - R_t, \beta P_{n+1}(q_t) - R_t) \\
&= \beta \max(\bar{p}_t, P_{n+1}(q_t)) - R_t \\
&= \beta \max(\hat{p}_t, P_n(q_t), P_{n+1}(q_t)) - R_t \\
&= \beta \max(\hat{p}_t, P_{n+1}(q_t)) - R_t
\end{aligned}$$

Therefore, as $n = T - t$, it is clear that (37) also holds for t. ■

Corollary 1. *The minimal worst-case regret $D(\beta)$ for the one-way trading problem is a convex function of β:*

$$D(\beta) = \beta(M - m)\left(1 - \frac{1}{\beta T}\right)^{+T} - (1 - \beta)m, \tag{44}$$

Proof: In the first period, there is $q_1 = 1, r_1 = 0, \hat{p}_1 = m$. Use these in (37) and simplify to have the result. The convexity of $D(\beta)$ is a consequence of the reward convexity in the one-way trading problem and Theorem 4. ■

Note that the result of Wang et al. (2016) is a special case of Theorem 7 with $\beta = 1$, and the proof for this general result uses quite different technical approach and tactics than those for the special case. Theorem 7 easily leads to a tremendously simplified derivation of the competitive ratio, as compared to the highly complicated analysis of El-Yaniv et al. (2001).

Corollary 2. *The competitive ratio defined in (16) for the one-way trading problem is the unique root β_0 of $D(\beta)$ as defined in (44).*

Proof: As $r^*(\zeta) \geq m > 0$, it follows from Lemma 4 and Lemma 5. ∎

This is in perfect agreement with El-Yaniv et al. (2001), except that they define competitive ratio as the inverse of β_0. Their analysis heavily relies on insights of the worst case price paths and is much more involved than this analysis, while this analysis can easily deduce all worst-case price paths in the same way as in Wang et al. (2016).

Corollary 3. *As β increases, the optimal trading policy becomes more aggressive by taking on more risk as it tends to reserve more inventory for the future and trade less in the current period with other things being equal:*

$$\forall h_t, \forall p_t, \pi_t^*(h_t, p_t; \beta_1) \leq \pi_t^*(h_t, p_t; \beta_2) \text{ if } \beta_1 > \beta_2 > 0.$$

Proof: Consider the quantity reserved for the future q_{t+1}^* in (38) and note that

$$P_{T-t}^-(p) = \beta(T-t)\left(1 - \sqrt[T-t]{\frac{p-m}{M-m}}\right)$$

increases in β, therefore q_{t+1}^* increases as β increases. ∎

Corollary 3 illustrates the continuous moderation of conservatism by β, and that the optimal policy gets more aggressive as β increases.
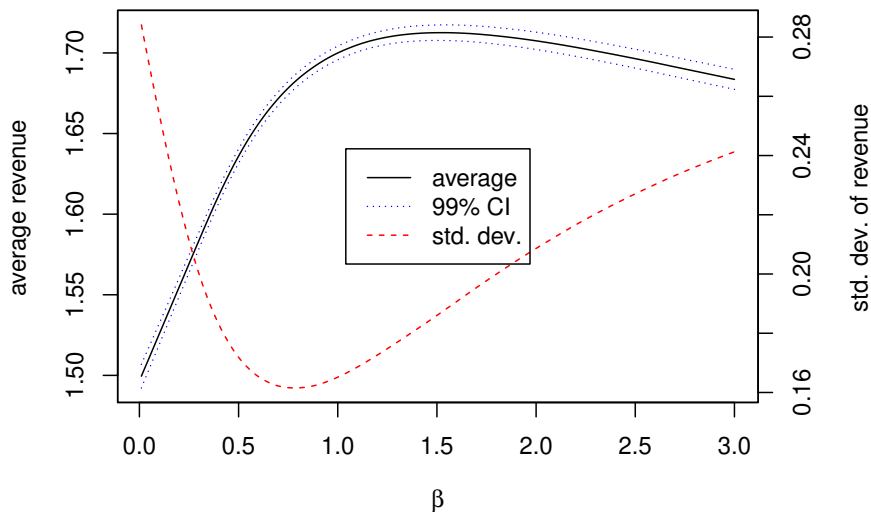
Figure 1: Average revenue and standard deviation of revenue for $T = 5$

5.3. Numerical Study.

The effect of β on the one-way trading policy is further demonstrated by numerical simulations. The prices are I.I.D. with a uniform distribution on $[1, 2]$ for all the $T = 5$ periods, but the ARM formulation only knows the price bounds $m = 1, M = 2$. For $\beta \in \{i/300 : i = 1, \cdots, 300\}$, the policy from (38) is executed on the same sequence of randomly generated prices and the overall revenue is accrued over all periods. Repeat this process $N = 10,000$ times and the average and standard deviation of the revenue for each β is calculated, together with a 99% confidence interval for the average revenue. The outcome is shown in Fig. 1.

Three distinct phases can be identified in Fig. 1 from left to right. The first phase ($\beta < 0.8$) witnesses increases in the average overall revenue with

decreases in the standard deviation as β increases, indicating that an overly conservative policy can not only hurt the revenue, but it may not really reduce the overall risk as supposed. To understand this apparent dilemma, note that the overall risk is different from the risk conditioned on historic prices: the former is estimated from runs that randomize all the prices, while the latter must fix the historic prices and only randomize the future prices. It is helpful to point out that more conservative policies with smaller β values indeed have lower conditional risks. But for the overall risk at $\beta = 0$, the policy trades everything in the first period at any price, while the policy at larger β diversifies the risks by trading less at lower prices and reserves more for future opportunities, which not only increases expected revenue but also helps reduce the overall risk. The second phase $(0.8 < \beta < 1.5)$ observes the normal case of increases in both revenue and overall risk. The third phase $(\beta > 1.5)$ experiences decreasing revenue while the risk increases. This is due to that an overly aggressive policy reserves too much quantity for the future and boldly takes the risk of selling a significant amount at whatever price in the last period.

This experiment provides some valuable insights, that both extreme conservatism $(\beta = 0)$ and extreme aggressiveness $(\beta \uparrow \infty)$ give poor performance with low expected reward and high overall risk, and the ARM criterion may trade off and find a sweet spot in between. An overly conservative robust policy by the maximin reward criterion (i.e. $\beta = 0$) can suffer both lower performance and higher overall risk. On the other hand, an overly aggressive robust policy for a big β may also hurt the performance as it boldly exposes to future price risks by reserving too much inventory, ending up selling it

in the last period at any price. Last but not least, there is a trade-off between the two extremes by adjusting the β value of the ARM criterion. As illustrated in Figure 1, the fine-tuned β may well go beyond $\beta = 1$ for the minimax absolute regret criterion.

6. Conclusion.

This paper proposes the ARM criterion with the β parameter for continuous control of conservatism. By minimizing the guarantee of worst-case regret against a β-adjusted benchmark that becomes more aggressive as β increases, the ARM criterion chooses a solution that is likely to "mimic" the benchmark's behavior, so that the conservatism of the recommended solution is moderated. Various theoretical properties of the ARM criterion are studied, such as continuity, monotonicity, and convexity, which may facilitate the analysis of problems, finding closed-form solutions, or designing better numerical algorithms to calculate competitive ratios. These theoretical investigations also lead to a new approach for competitive ratio analysis, which may be much simpler than the traditional approach for some problems, such as observed in the analysis of the one-way trading problem.

The tractability is studied for two-stage linear problems with the ARM criterion. Two particular situations are studied: the right-hand side uncertainty and the objective uncertainty. Equivalent reformulations into TSLRO/FR problems make it possible to take advantage of the tractable solution methods developed recently to solve them for practical applications.

Finally, the ARM criterion is applied to the robust one-way trading problem to demonstrate its properties and potential. Closed-form solutions are

obtained, which facilitates the derivation of competitive ratios by the new approach. The optimal policy is shown to get more aggressive as the control parameter increases. A numerical study is carried out to demonstrate the effects of smooth control of conservatism for the one-way trading problem. Some insights are gleaned from the numerical study: extreme conservatism or aggressiveness may suffer from both lower average reward and higher overall risk at the same time, and the proper β value is found somewhere in between.

The investigations of the ARM criterion carried out in this paper may serve as a starting point for future research. A rigorous analysis on how β moderates conservatism is theoretically interesting and challenging at the same time. Conceivably, how to choose an appropriate β value in real applications depends on the application context, and it is another worthy topic of interest for future study.

References

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

Averbakh, I. (2005). Computing and minimizing the relative regret in combinatorial optimization with interval data. *Discrete Optimization*, 2(4):273–287.

Averbakh, I. and Lebedev, V. (2005). On the complexity of minmax regret linear programming. *European Journal of Operational Research*, 160(1):227–231.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). *Robust optimization.* Princeton university press.

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

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research*, 23(4):769–805.

Ben-Tal, A. and Nemirovski, A. (1999). Robust solutions of uncertain linear programs. *Operations research letters*, 25(1):1–13.

Ben-Tal, A. and Nemirovski, A. (2000). Robust solutions of linear programming problems contaminated with uncertain data. *Mathematical programming*, 88(3):411–424.

Ben-Tal, A. and Nemirovski, A. (2008). Selected topics in robust convex optimization. *Mathematical Programming*, 112(1):125–158.

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

Bertsimas, D., Iancu, D. A., and Parrilo, P. A. (2010). Optimality of affine policies in multistage robust optimization. *Mathematics of Operations Research*, 35(2):363–394.

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

Borodin, A. and El-Yaniv, R. (2005). *Online computation and competitive analysis.* cambridge university press.

Caldentey, R., Liu, Y., and Lobel, I. (2017). Intertemporal pricing under minimax regret. *Operations Research*, 65(1):104–129.

Delage, E. and Iancu, D. A. (2015). Robust multistage decision making. In *The operations research revolution*, pages 20–46. INFORMS.

El-Ghaoui, L. and Lebret, H. (1997). Robust solutions to least-square problems to uncertain data matrices. *Sima Journal on Matrix Analysis and Applications*, 18:1035–1064.

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

El-Yaniv, R., Fiat, A., Karp, R. M., and Turpin, G. (2001). Optimal search and one-way trading online algorithms. *Algorithmica*, 30(1):101–139.

Gabrel, V., Murat, C., and Thiele, A. (2014). Recent advances in robust optimization: An overview. *European journal of operational research*, 235(3):471–483.

Kouvelis, P. and Yu, G. (2013). *Robust discrete optimization and its applications*, volume 14. Springer Science & Business Media.

Lan, Y., Gao, H., Ball, M. O., and Karaesmen, I. (2008). Revenue management with limited demand information. *Management Science*, 54(9):1594–1609.

Megiddo, N. (1978). Combinatorial optimization with rational objective functions. In *Proceedings of the tenth annual ACM symposium on Theory of computing*, pages 1–12.

Milnor, J. (1951). Games against nature. Technical report, RAND PROJECT AIR FORCE SANTA MONICA CA.

Natarajan, K., Shi, D., and Toh, K.-C. (2014). A probabilistic model for minmax regret in combinatorial optimization. *Operations Research*, 62(1):160–181.

Özkaya, M., İzgi, B., and Perc, M. (2022). Axioms of decision criteria for 3d matrix games and their applications. *Mathematics*, 10(23):4524.

Perakis, G. and Roels, G. (2008). Regret in the newsvendor model with partial information. *Operations research*, 56(1):188–203.

Poursoltani, M. and Delage, E. (2021). Adjustable robust optimization reformulations of two-stage worst-case regret minimization problems. *Operations Research*.

Savage, L. J. (1951). The theory of statistical decision. *Journal of the American Statistical association*, 46(253):55–67.

Shapiro, A. (2017). Interchangeability principle and dynamic equations in risk averse stochastic programming. *Operations Research Letters*, 45(4):377–381.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints

and applications to inexact linear programming. *Operations research*, 21(5):1154–1157.

Stoye, J. (2011). Statistical decisions under ambiguity. *Theory and decision*, 70(2):129–148.

Vinod, B. (2021). An approach to adaptive robust revenue management with continuous demand management in a covid-19 era. *Journal of Revenue and Pricing management*, 20(1):10–14.

Wang, W., Wang, L., Lan, Y., and Zhang, J. X. (2016). Competitive difference analysis of the one-way trading problem with limited information. *European Journal of Operational Research*, 252(3):879–887.

Yanıkoğlu, İ., Gorissen, B. L., and den Hertog, D. (2019). A survey of adjustable robust optimization. *European Journal of Operational Research*, 277(3):799–813.

Zeng, B. and Zhao, L. (2013). Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5):457–461.