PHBS WORKING PAPER SERIES

Fine Control of Conservatism for Robust Optimization by Adjustable Regret

Yingjie Lan
Peking University

September 2025

Working Paper 20250902

Abstract

Overconservatism has long been recognized as a major issue of robust optimization, despite its major advantages of tractability, performance guarantee, and limited information. A new criterion based on adjustable regret is proposed to address this issue by adapting the level of conservatism to the environment, while maintaining all the aforementioned advantages. The level of conservatism can be fine-tuned by maximizing the reward guarantee for scenarios representative of opportunities provided by experts as most likely values, leading to a simple heuristic to best catch opportunities. This criterion also supports a new approach to competitive ratio analysis that is applicable even to multistage problems. The new criterion is then applied to the one-way trading problem with analytical solutions, from which the competitive ratio is easily derived by the new approach. Numerical experiments are conducted to demonstrate fine control of conservatism and the effectiveness of the heuristic, with the average reward improved in one case by 3 - 9% over other commonly used criteria.

Keywords: robust optimization, decision criteria, overconservatism, convex conjugate, one-way trading
JEL Classification: C44, C61, C63

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

Fine 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

Overconservatism has long been recognized as a major issue of robust optimization, despite its major advantages of tractability, performance guarantee, and limited information. A new criterion based on adjustable regret is proposed to address this issue by adapting the level of conservatism to the environment, while maintaining all the aforementioned advantages. The level of conservatism can be fine-tuned by maximizing the reward guarantee for scenarios representative of opportunities provided by experts as most likely values, leading to a simple heuristic to best catch opportunities. This criterion also supports a new approach to competitive ratio analysis that is applicable even to multistage problems. The new criterion is then applied to the one-way trading problem with analytical solutions, from which the competitive ratio is easily derived by the new approach. Numerical experiments are conducted to demonstrate fine control of conservatism and the effectiveness of the heuristic, with the average reward improved in one case by 3 - 9% over other commonly used criteria.

Keywords: robust optimization, decision criteria, overconservatism, convex conjugate, one-way trading

1. Introduction.

Robust optimization (RO) is a popular method of decision making under uncertainty, and the decision criterion plays a key role in achieving the major advantages of RO: limited informatin requirement, robust solutions with a performance guarantee, and computational tractability. Limited information for RO is provided by an uncertainty set with all possible scenarios, without any probabilistic distribution as required by stochastic optimization. The outcome of an action depends on the realized scenario, it relies on the criterion to evaluate an action when only the outcomes for scenarios are known with limited information. The recommended solutions are robust as the criterion endows them with a performance guarantee for worst scenarios. And finally, computational tractability generally relies on the criterion preserving convexity, as many classes of convex optimization problems admit polynomial-time algorithms (Nesterov and Nemirovskii 1994), while mathematical optimization is NP-hard in general (Murty and Kabadi 1987). The convexity of the feasible region is naturally preserved in RO, so convexity is preserved as long as the criterion forms a convex objective. These advantages have made RO an appealing method in various fields of applications, such as finance, operations management, aerospace, and defense.

Despite these advantages, a major issue with RO is overconservatism, which has led to flurries of highly fruitful researches. Obsessed with the worst scenario while ignoring all opportunities in others, no matter how likely they may occur, overconservatism ends up sacrificing too much performance for the sake of robustness, which seriously hinders the adoption of RO in industries, such as robust revenue management in airlines (Vinod 2021). Such an

obsession is abated by having some ambiguous information on probability distributions in distributionally robust optimization (DRO), a less conservative method that has attracted great research interests, see Kuhn et al. (2025) for a recent survey and Zhen et al. (2025) for a unified thoery. Meanwhile, combating overconservatism without distributions has always been an active and important research front for RO soon after the first RO models appear in Soyster (1973). These early models are extremely pessimistic on how uncertainty affects feasibility and performance, contributing to overconservatism in two ways: Firstly, extreme scenarios may render some good solutions infeasible; Secondly, the minimax (cost) criterion evaluates a solution by the worst scenario while ignoring all opportunities in others. Hence overconservatism has been tackled accordingly by either excluding certain extreme scenarios or resorting to less conservative criteria.

Scenario exclusion begins by questioning the absolute guarantee of feasibility under all scenarios, especially those rare and extreme ones. Though necessary for critical applications where infeasibility causes disasters like doomed satellites or damaged rovers, it can be relaxed if adverse events only bring about limited consequences, such as low demand or supply in business. In the latter case, it is acceptable to exclude some rare and extreme scenarios and have a smaller uncertainty set, settling for a probabilistic feasibility guarantee in exchange for better performance. Researches in this regard provide insights into robust solutions as well as probabilistic guarantees of feasibility, see Ben-Tal and Nemirovski (2008) and Bertsimas et al. (2011) for a comprehensive survey and Ben-Tal et al. (2009) for a book treatment. Models with ellipsoidal uncertainty sets are proposed 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 computationally demanding, Bertsimas and Sim (2004) proposes uncertainty budget to fully control the uncertainty set while maintaining linearity. In this approach, the uncertainty set is tampered with to trade-off between robustness and performance, which can be difficult in practice as it may need probability estimation for those rare and extreme scenarios.

Another direction to tackle overconservatism is by employing alternative decision criterion. The minimax criterion adopted in the earliest RO models is invented when economists contemplated on decision theories (Wald 1950). This criterion is critisized by Savage (1954) as "ultrapessimistic", as it focuses entirely on performances in the worst scenario and ignores all plausible opportunities in others. The opportunity in a scenario is fully realized in and measured by the ex post (i.e. after knowing the scenario) optimal objective. The regret of a solution is the opportunity loss defined as the difference between the objectives of the ex post optimal and the solution. Savage (1951) proposes the minimax regret criterion as a less conservative criterion that minimizes the worst-case regret, which also serves as a *regret guarantee*. The so-called "competitive ratio," a popular criterion for online optimization (Borodin and El-Yaniv 2005; Kouvelis and Yu 2013), is equivalent to the relative regret criterion, which considers the ratio of regret to the ex post optimal objective instead. Various numerical studies find the absolute regret less conservative than the relative regret, which is in turn less conservative than the maximin criterion in reward maximizing problems (Lan et al. 2008; Perakis and Roels 2008; Poursoltani and Delage 2021).

All these commonly used criteria maintain the major advantages of RO, yet they only offer three distinct levels of conservatism, and it is only possible with a continuum of choices to have fine control of conservatism. An early attempt is made by Hurwicz (1951) with the Hurwicz criterion, which evaluates a solution by a weighted sum of its worst and best outcomes in order to balance pessimism and optimism. There is a "coefficient of pessimism" (α) between 0 and 1 as the weight on the worst outcome, while $(1-\alpha)$ weights on the best outcome. It becomes the most conservative as the minimax criterion when $\alpha = 1$, and the most aggressive as the minimin criterion when $\alpha = 0$, and in between it generally gets more and more conservative as α increases. Unfortunately, its major drawback is not preserving convexity, thus losing the advantage of computational tractability.

Many other criteria are proposed thereafter to moderate conservatism for fine control, but none of them can keep all the major advantages of RO. For example, the p-robustness by Snyder (2006) first screens out by additional constraints overly conservative solutions whose worst-case regret exceeds an upper limit, but sometimes it is very difficult to determine if it is making a problem infeasible. Kalaı et al. (2012) suggest the lexicographic robustness criterion to mitigate the primary role of the worst-case scenario in solution evaluation, yet it requires finite uncertainty sets, and poses a serious computational challenge for large uncertainty sets.

This work aspires to create a new criterion for fine control of conservatism while upholding all the major advantages of RO. Adjustable regret minimization (ARM) is proposed as a new criterion with a continuum of conseratism choices. The regret becomes adjustable by comparing the actual

objective not with a fixed ex post optimal, but with that ex post optimal scaled by a *conservatism control parameter* (CCP). It turns out this simple design works with limited information, comes with performance guarantees, and maintains tractability by preserving convexity. Most interestingly, the criterion generally gets more aggressive as the CCP increases, which enables the CCP to adjust the level of conservatism for fine control. By optimizing certain performance measures, such as the reward guarantee for scenarios representative of opportunities, derived from experts estimating most likely values (as for task durations in project management), a simple heuristic to determine the CCP is proposed, analyzed theoretically, and studied numerically with impressive results. The application of the ARM criterion and the heuristic is not limited to situations without distributions. When distribution is available but a performance guarantee is highly desireable, they can help find a robust solution with maximal expected reward. They can also be applied with DRO to queeze out extra performance if the most likely subset of distributions are known, which is often the case when confidence intervals of distribution parameters are estimated to specify the ambiguity set for DRO.

Many of the nice analytical properties of the ARM criterion support a new approach to competitive ratio analysis that may significantly reduce the analysis complexity to derive closed-form solutions. Competitive ratio analysis is often more complex than absolute regret analysis, which can be observed in comparing El-Yaniv et al. (2001) and Wang et al. (2016), as each carries out one type of analysis with exactly the same problem setup. The new approach first solves the problem with the ARM criterion, which can

have similar complexity as the absolute regret analysis, then the competitive ratio is derived simply by solving an equation. This new approach is demonstrated by solving the robust one-way trading problem, recovering the main results in both papers.

The contributions of this paper are as follows: (i) The ARM criterion is proposed for fine control of conservatism while the major advantages of RO are maintained. The properties of the ARM criterion are studied theoretically, such as convexity preservation and conservatism control. (ii) The mechanism to control conservatism is investigated and a heuristic is proposed to determine the CCP by maximizing the reward guarantee for representative scenarios, while alternatives are also discussed. (iii) A new approach to competitive ratio analysis based on the ARM criterion is investigated, which can reduce the complexity of analysis to find analytical solutions. (iv) The robust one-way trading problem with the ARM criterion is solved analytically, with the competitive ratio easily derived by the new approach, and numerical experiments are conducted to demonstrate fine control of conservatism, with the average reward of the heuristic witnessing a 3% to 9% improvement in one case over other commonly used criteria.

The rest of this paper is arranged as follows. Section 2 gives general ARM formulations, whose properties are studied in Section 3, leading to a new approach to competitive ratio analysis. The mechanism for fine control of conservatism is discussed and analyzed with a heuristic to determine a proper CCP. In section 4 the ARM criterion is applied to the robust one-way trading problem. The analysis derives a closed-form solution, which yields the competitive ratio quickly by the new appraoch. Numerical experiments

demonstrate effective fine control of conservatism by the ARM criterion and the heuristic. Finally, section 5 draws conclusions with future research outlooks.

2. Formulations.

The ARM criterion is first presented with single-stage problems for clarity, then extended to multistage formulations for generality. Though only reward maximization is considered here, the results should translate to cost minimization. A scenario ζ consists of realized values for uncertain data, and all scenarios is specified by the uncertainty set $\mathcal{U}$. Let X_ζ be the feasible set for ζ, as the constraints may depend on ζ. The robustly feasible set $X = \bigcap_{\zeta \in \mathcal{U}} X_\zeta$ is feasible for all scenarios. For generality, both $\mathcal{U}$ and X_ζ can be continuous or discrete. In single-stage problems, an action x takes place first, then a scenario ζ realizes, and the resultant reward $r(x, \zeta)$ depends on both x and ζ. Let $r^*(\zeta) = \max_{x \in X_\zeta} r(x, \zeta)$ denote the ex post optimal, measuring the potential opportunities in ζ. It is assumed throughout the paper that the min and max operators are well-defined, otherwise they may be replaced by inf and sup.

The CCP $\beta \in [-\infty, +\infty]$ is introduced into ARM as a factor in the benchmark $\beta r^*(\zeta)$ for the actual reward $r(x, \zeta)$, producing a β-adjusted regret $D(x, \zeta; \beta) = \beta r^*(\zeta) - r(x, \zeta)$ for an action $x \in X$. The worst-case regret $\bar{D}(x; \beta) = \max_{\zeta \in \mathcal{U}} D(x, \zeta; \beta)$ serves as a regret guarantee for x. The ARM criterion minimizes this regret guarantee:

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

Though (1) is used for analyzing and reducing computational complexities in combinatorial optimization problems in Averbakh (2005), it has never been proposed and studied as a new criterion for moderating conservatism. Let $\check{r}(r, \beta) = \beta r - D(\beta)$ denote the *reward guarantee* by the optimal solution for any $\zeta \in U(r) = \{\zeta \in \mathcal{U} : r^*(\zeta) = r\}$. Obviously, the ARM criterion chooses the same solutions if $r(x, \zeta)$ is replaced by $r'(x, \zeta) = kr(x, \zeta) + b$ for any $k > 0$ and $b \in \mathbb{R}$. Also note that it can easily adapt to DRO with ζ representing a distribution and $r(x, \zeta)$ being the expected revenue.

Here is a glimpse of the effects of β on conservatism when the ARM criterion transmorphs into other well-known criteria. At $\beta = 0$ it becomes the maximin criterion in traditional RO models, which is the most conservative. When β takes on the competitive ratio (usually a special value between 0 and 1), it is the same as the relative regret criterion (more details later), which is less conservative than the previous. At $\beta = 1$ it is the absolute regret criterion that is even less conservative, and finally as $\beta = +\infty$ it recovers the maximax criterion that is the most aggressive. Interestingly, at the other extreme of $\beta = -\infty$ is the opposite to the maximax criterion, which recommends an x that performs best for the least promising scenario ζ' with $r^*(\zeta')$ being the minimal. These cases suggests that the ARM criterion becomes increasingly more aggressive as β gets bigger, which will be analyzed in depth later.

The formulation readily extends to multistage problems, where decisions are made stage by stage as the scenario gradually reveals itself. Let $t = 1, \cdots, T$ labels the stages sequentially, and let x_t and ζ_t denote the part of decision and scenario in stage t. To standardize and simplify, a stage decision x_t is always carried out before the stage scenario ζ_t is realized and known,

which is without loss of generality: If a stage scenario is realized before any decision is made, a dummy decision can be inserted in the very beginning for a standardized formulation. Let $x = (x_1, \cdots, x_T)$ and $\zeta = (\zeta_1, \cdots, \zeta_T)$ for the entire decision and scenario.

It is assumed that decision makers neither know nor influence the stage scenarios not yet realized when making a stage decision, which is known in multistage stochastic programming as *nonanticipativity*. Specifically, when making decision x_t for stage t, only the partial scenario $\zeta_{1:t} = (\zeta_1, \cdots, \zeta_{t-1})$ is known, which reduces the uncertainty set to $\mathcal{U}(\zeta_{1:t}) = \{\omega \in \mathcal{U} : \omega_{1:t} = \zeta_{1:t}\}$. For the current $\mathcal{U}(\zeta_{1:t})$, let $X_{\mathcal{U}(\zeta_{1:t})} = \bigcap_{\omega \in \mathcal{U}(\zeta_{1:t})} X_\omega$ be the feasible set, which expands as uncertainty reduces: $X_{\mathcal{U}(\zeta_{1:t})} \subseteq X_{\mathcal{U}(\zeta_{1:t+1})}$ as $\mathcal{U}(\zeta_{1:t+1}) \subseteq \mathcal{U}(\zeta_{1:t})$. Let $x_{1:t} = (x_1, \cdots, x_{t-1})$ be all stage decisions before stage t, and $h_t = (x_{1:t}, \zeta_{1:t})$ be the history. Given $h_t = (x_{1:t}, \zeta_{1:t})$, the compatible feasible set is $X(h_t) = \{y \in X_{\mathcal{U}(\zeta_{1:t})} : y_{1:t} = x_{1:t}\}$. Let $X_t(h_t) = \{y_t : y \in X(h_t)\}$ be the stage feasible set, and $\mathcal{U}_t(\zeta_{1:t}) = \{\omega_t : \omega \in \mathcal{U}(\zeta_{1:t})\}$ be the stage scenario set. A history $h_t = (x_{1:t}, \zeta_{1:t})$ is feasible only if it satisfies

$$x_\tau \in X_\tau(h_\tau), \tau = 1, \cdots, t-1, \text{ where } h_\tau = (x_{1:\tau}, \zeta_{1:\tau}). \tag{2}$$

Let H_t denote the set of all feasible histories before stage t.

The ARM criterion for multistage problems can be defined recursively. Let $r(x, \zeta)$ denote the total reward over all stages, either accrued over stages or received at once in the end. Let $r^*(\zeta) = \max_{x \in X_\zeta} r(x, \zeta)$ be the ex post optimal. The regret for a completed history $h_{T+1} = (x, \zeta)$ is

$$D_T(h_{T+1}; \beta) = \beta r^*(\zeta) - r(x, \zeta). \tag{3}$$

Work from $D_T(h_{T+1}; \beta)$, the regret guarantee $\bar{D}_t(x_t, h_t; \beta)$ and the minimum

guarantee $D_{t-1}(h_t;\beta)$ is found by *backward induction* for $t = T, \cdots, 1$:

$$\begin{aligned} \bar{D}_t(x_t, h_t;\beta) &= \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t(h_{t+1};\beta), & (4)\\ D_{t-1}(h_t;\beta) &= \min_{x_t \in X_t(h_t)} \bar{D}_t(x_t, h_t;\beta), \\ &= \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t(h_{t+1};\beta). & (5) \end{aligned}$$

where h_{t+1} is formed by appending x_t and ζ_t to $x_{1:t}$ and $\zeta_{1:t}$ respectively. Note that (4) is often referred to as the "adversarial problem", as if an almighty adversary is making the worst for the decision maker. As h_1 is empty, let $D(\beta) \equiv D_0(h_1;\beta)$ for the best overall regret guarantee, so that there is still $\check{r}(r,\beta) = \beta r - D(\beta)$. This completes the *plain formulation*.

An alternative formulation is based on policies. A feasible policy π is a sequence of functions $\pi = \{\pi_t : \pi_t(h_t) \in X_t(h_t), \forall h_t \in H_t, t = 1, 2, \cdots, T\}$ to make stage decisions by $x_t = \pi_t(h_t)$, with nonanticipativity already baked in. The focus is on deterministic policies, though random ones are possile. A policy π can determine the whole decision x for a scenario ζ, or simply $x = \pi(\zeta)$, and the subset of realizable histories before stage t under π is $H_t^\pi = \{(x_{1:t}, \zeta_{1:t}) : x = \pi(\zeta), \zeta \in \mathcal{U}\}$. The restriction of π_t to H_t^π for all t defines a pruned policy $\ddot{\pi}$, as other histories are cut away. Let Π be the set of all feasible deterministic policies, and let $\ddot{\Pi} = \{\ddot{\pi} : \pi \in \Pi\}$.

The regret guarantee of π is also found by backward induction. In the end the history $h_{T+1} = (x, \zeta)$ is fully developed, and the regret is still

$$D_T^\pi(h_{T+1};\beta) = \beta r^*(\zeta) - r(x,\zeta). \quad (6)$$

The regret guarantee given $h_t = (x_{1:t}, \zeta_{1:t})$ is defined recursively by

$$D_{t-1}^\pi(h_t;\beta) = \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t^\pi(h_{t+1}^\pi;\beta),\ t = T, \cdots, 1, \quad (7)$$

where $h^\pi_{t+1} = ((x_{1:t}, \pi_t(h_t)), (\zeta_{1:t}, \zeta_t))$ evolves from h_t under π. Apply (7) recursively for the overall regret guarantee $D^\pi(\beta) \equiv D^\pi_0(h_1, \beta)$:

$$\begin{aligned} D^\pi(\beta) &= \max_{\zeta_1 \in \mathcal{U}_1(\zeta_{1:1})} D^\pi_1(h^\pi_2; \beta) \\ &= \max_{\zeta_1 \in \mathcal{U}_1(\zeta_{1:1})} \max_{\zeta_2 \in \mathcal{U}_2(\zeta_{1:2})} D^\pi_2(h^\pi_3; \beta) \\ &= \max_{\zeta_1 \in \mathcal{U}_1(\zeta_{1:1})} \cdots \max_{\zeta_T \in \mathcal{U}_T(\zeta_{1:T})} D^\pi_T(h^\pi_{T+1}; \beta) \\ &= \max_{\zeta \in \mathcal{U}} D^\pi_T(h^\pi_{T+1}; \beta) \end{aligned} \tag{8}$$

where $h^\pi_{T+1} = (\pi(\zeta), \zeta)$. The *policy-based formulation* chooses a policy to minimize the overall regret guarantee:

$$\min_{\pi \in \Pi} D^\pi(\beta) = \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r(\pi, \zeta), \tag{9}$$

where $r(\pi, \zeta) = r(\pi(\zeta), \zeta)$, as π in (9) plays the same role as x in (1).

A couple of comments on the policy-based formulation. First, in (9) all decisions are made by a policy π, reaching only realizable histories in H^π_t, thus it makes no difference if Π is replaced by $\ddot{\Pi}$ there. Second, for problems with states, any $h_t \in H_t$ maps to a state $s(h_t)$, which can replace h_t to reduce complexity. Finally, there may be many policies that give the same $D(\beta)$, but have different values of $D^\pi_{t-1}(h_t; \beta)$ for some history h_t. The optimal policies can be refined by the principle of optimality (Bellman 1954), which requires recursively that the subpolicies of an optimal policy are themesleves optimal: A policy π^* is optimal if and only if $D^{\pi^*}_{t-1}(h_t; \beta) \leq D^\pi_{t-1}(h_t; \beta)$ for all $\pi \in \Pi$ and $h_t \in H_t, t = 1, \cdots, T$.

3. Theoretical Analysis.

In this section, formulation equivalence and convexity preservation are established first, then comes a new approach to competitive ratio analysis,

and finally fine control of conservatism is analyzed with a heuristic proposed.

Theorem 1. *Under the principal of optimality, the plain and the policy-based formulations are equivalent in that any optimal policy* $\pi^* \in \Pi$ *satisfies*

$$D_{t-1}(h_t;\beta) = D^{\pi^*}_{t-1}(h_t;\beta), \forall h_t \in H_t, t = 1, \cdots, T+1, \tag{10}$$

$$\pi^*_t(h_t) \in \underset{x_t \in X_t(h_t)}{\operatorname{argmin}} \max_{\zeta_t \in \mathcal{U}_t(\zeta_{1:t})} D_t(h_{t+1};\beta), \forall h_t \in H_t, t = 1, \cdots, T, \tag{11}$$

where $\zeta_{1:t}$ *belongs to* $h_t = (x_{1:t}, \zeta_{1:t})$.

Proof: A policy $\pi^* \in \Pi$ satisfying (11) clearly exists for well-defined problems, and its optimality can be proven in two logic progressions. The first progression proves that if any policy π^* satisfies (11), then it also satisfies (10), which is done by backward induction. As the initial step, (10) trivially holds for $t = T+1$. For the induction step, assume (10) holds for $t = \tau+1$, and show it also holds for $t = \tau$. Recall (5) and proceed as follows

$$\begin{aligned} D_{\tau-1}(h_\tau;\beta) &= \min_{x_\tau \in X_\tau(h_\tau)} \max_{\zeta_\tau \in \mathcal{U}_\tau(\zeta_{1:\tau})} D_\tau(h_{\tau+1};\beta) \\ &= \max_{\zeta_\tau \in \mathcal{U}_\tau(\zeta_{1:\tau})} D_\tau(h_{\tau+1};\beta) \\ &= \max_{\zeta_\tau \in \mathcal{U}_\tau(\zeta_{1:\tau})} D^{\pi^*}_\tau(h^{\pi^*}_{\tau+1};\beta) \\ &= D^{\pi^*}_{\tau-1}(h_\tau;\beta), \end{aligned}$$

where the second equality comes by (11), the third by the inductive assumption, and the last comes by (7). Thus π^* also satisfies (10).

The next progression proves that if a policy π^* satisfies (10), then it is optimal under the principle of optimality, which becomes

$$D_{t-1}(h_t;\beta) \le D^{\pi}_{t-1}(h_t;\beta), \forall \pi \in \Pi, \forall h_t \in H_t, t = 1, \cdots, T+1, \tag{12}$$

after integrating (10). It is again by backward induction, whose initial step trivially holds for $t = T + 1$. For the induction step, assume (12) holds for $t = \tau + 1$ so as to show it also holds for $t = \tau$. Recall (5) with $t = \tau$ and apply the assumption by replacing $D_\tau(h_{\tau+1}; \beta)$ with $D^\pi_\tau(h_{\tau+1}; \beta)$:

$$\begin{aligned} D_{\tau-1}(h_\tau; \beta) &\leq \min_{x_\tau \in X_\tau(h_\tau)} \max_{\zeta_\tau \in \mathcal{U}_\tau(\zeta_{1:\tau})} D^\pi_\tau(h_{\tau+1}; \beta) \\ &\leq \max_{\zeta_\tau \in \mathcal{U}_\tau(\zeta_{1:\tau})} D^\pi_\tau(h^\pi_{\tau+1}; \beta) \\ &= D^\pi_{\tau-1}(h_\tau; \beta), \end{aligned}$$

where the second inequality comes by fixing $x_\tau = \pi_\tau(h_\tau)$, and the last line comes by (7). Therefore (12) holds by backward induction, and π^* is optimal.

Once an optimal policy π^* is known to satisfy (10), the principle of optimality requires that any optimal policy must satisfy (10), thus (11) automatically holds. ■

It is handy to have both formulations, as the plain one solves the problem stage by stage, while the policy-based one facilitates theoretical analysis. The interchangeability principle of Shapiro (2017) can obtain a weaker result, as it does not respect the principle of optimality. Also note that with $t = 1$ there is $D^{\pi^*}(\beta) = D(\beta)$ by (10).

Convexity preservation is crucial for compuational tractability, which works with consistently convex problems and pruned policies. Consistent convexity requires that $\max\{r(x, \zeta) : x \in X_\zeta\}$ is convex for any $\zeta \in \mathcal{U}$. A convex combination of two pruned policies $\ddot{\pi}_1, \ddot{\pi}_2 \in \ddot{\Pi}$ by a $\lambda \in [0, 1]$ is defined as $\ddot{\pi}(\zeta) = \lambda\ddot{\pi}_1(\zeta) + (1 - \lambda)\ddot{\pi}_2(\zeta)$ for all $\zeta \in \mathcal{U}$, which can be written as $\ddot{\pi} = \lambda\ddot{\pi}_1 + (1 - \lambda)\ddot{\pi}_2$ with $\ddot{\pi}$ regarded as a vector of $\ddot{\pi}(\zeta), \zeta \in \mathcal{U}$.

Theorem 2. *With consistent convexity, the policy-based formulation* (9) *restricted to pruned policies* $\ddot{\Pi}$ *is convex, and all subproblems* (5) *in the plain formulation are convex.*

Proof: First prove the convexity of the domains. As a projection of the convex set $X(h_t)$, the convexity of $X_t(h_t)$ is obvious, but the convexity of $\ddot{\Pi}$ needs some explanation. For $\ddot{\pi}^1, \ddot{\pi}^2 \in \ddot{\Pi}$, the histories $h_{T+1}^{\ddot{\pi}^i} = (\ddot{\pi}(\zeta), \zeta), i = 1, 2$ satisfy the feasibility condition (2) for all $\zeta \in \mathcal{U}$. For $\ddot{\Pi}$ to be convex, it must be shown that policy $\ddot{\pi} = \lambda\ddot{\pi}^1 + (1-\lambda)\ddot{\pi}^2$ for any $\lambda \in [0, 1]$ also satisfies (2) to have $\ddot{\pi} \in \ddot{\Pi}$. Apply both $\ddot{\pi}^i, i = 1, 2$ to an arbitrary ζ to have $x^i = \ddot{\pi}^i(\zeta), i = 1, 2$. As $h_{T+1}^{\ddot{\pi}^i} = (x^i, \zeta), i = 1, 2$ satisfy (2), there exists $y^{it} \in X_{\mathcal{U}(\zeta_{1:t})}$ such that $y_{1:t}^{it} = x_{1:t}^i$ for $i = 1, 2$ and $t = 1, \cdots, T$. Consistent convexity implies that any $X_{\mathcal{U}(\zeta_{1:t})}$ is convex, which ensures $y^t = \lambda y^{1t} + (1-\lambda)y^{2t} \in X_{\mathcal{U}(\zeta_{1:t})}$, hence for $x = \ddot{\pi}(\zeta), h_{T+1}^{\ddot{\pi}} = (x, \zeta)$ there is $x_t \in X_t(h_t^{\ddot{\pi}})$ for $t = 1, \cdots, T$, so $h_{T+1}^{\ddot{\pi}}$ satisfies (2) and there is $\ddot{\pi} \in \ddot{\Pi}$.

Next show the convexity of the objectives. The regret of any π with a given $\zeta \in \mathcal{U}$ is $E_\zeta(\pi) \equiv \beta r^*(\zeta) - r(\pi, \zeta)$. For $\ddot{\pi}_1, \ddot{\pi}_2 \in \ddot{\Pi}$ and a $\lambda \in [0, 1]$, let $\ddot{\pi} = \lambda\ddot{\pi}_1 + (1-\lambda)\ddot{\pi}_2 \in \ddot{\Pi}$. For any $\zeta \in \mathcal{U}$, there is $\ddot{\pi}(\zeta) = \lambda\ddot{\pi}_1(\zeta) + (1-\lambda)\ddot{\pi}_2(\zeta)$, thus $E_\zeta(\ddot{\pi}) \le \lambda E_\zeta(\ddot{\pi}_1) + (1-\lambda)E_\zeta(\ddot{\pi}_2)$ as $r(x, \zeta)$ is concave, so $E_\zeta(\pi)$ is convex on $\ddot{\Pi}$. The objective at $\ddot{\pi} \in \ddot{\Pi}$ in (9) is $D^{\ddot{\pi}}(\beta) = \max_{\zeta \in \mathcal{U}} E_\zeta(\ddot{\pi})$, which is convex in $\ddot{\pi}$ as a pointwise max of convex functions on $\ddot{\Pi}$.

It remains to show that $\bar{D}_t(x_t, h_t; \beta)$ in (5) is convex in x_t for any $h_t = (x_{1:t}, \zeta_{1:t}) \in H_t$. Let $\ddot{\Pi}(h_t) = \{\ddot{\pi} \in \ddot{\Pi} : \ddot{\pi}_\tau(\zeta_{1:\tau}) = x_\tau, \tau = 1, \cdots, t-1\}$ and $\ddot{\Pi}(x_t, h_t) = \{\ddot{\pi} \in \ddot{\Pi}(h_t) : \ddot{\pi}_t(\zeta_{1:t}) = x_t\}$, so that $\ddot{\Pi}(h_t) = \bigcup_{x_t \in X_t(h_t)} \ddot{\Pi}(x_t, h_t)$. Due to nonanticipativity, any $\ddot{\pi} \in \ddot{\Pi}(h_t)$ satisfies $\ddot{\pi}_t(\zeta_1') = \ddot{\pi}_t(\zeta_2')$ for any $\zeta_1', \zeta_2' \in \mathcal{U}(\zeta_{1:t})$, thus $\ddot{\pi}$ can be represented as a vector with only one x_t com-

ponent instead of many copies of the same x_t for each $\zeta' \in \mathcal{U}(\zeta_{1:t})$. Both $\ddot{\Pi}(x_t, h_t)$ and $\ddot{\Pi}(h_t)$ are convex as slices of $\ddot{\Pi}$, and $\ddot{\Pi}(x_t, h_t)$ is a slice of $\ddot{\Pi}(h_t)$. Apply Theorem 1 to the subproblem (4) as an independent problem with an uncertainty set $\mathcal{U}(\zeta_{1:t})$ and a dummy decision x_t so that $\ddot{\Pi}(x_t, h_t)$ has all pruned policies for it, to have $\bar{D}_t(x_t, h_t; \beta) = \min_{\ddot{\pi} \in \ddot{\Pi}(x_t, h_t)} g(\ddot{\pi})$, where $g(\ddot{\pi}) \equiv \max_{\zeta \in \mathcal{U}(\zeta_{1:t})} E_\zeta(\ddot{\pi})$ is convex over $\ddot{\Pi}(x_t, h_t)$. The epigraph $\mathbf{epi}\ \bar{D}_t(X_t(h_t), h_t; \beta) = \{(x_t, v) : (\ddot{\pi}, v) \in \mathbf{epi}\ g(\ddot{\Pi}(h_t)) \text{ for some } \ddot{\pi} \in \ddot{\Pi}(x_t, h_t)\}$ is convex as a projection of the convex set $\mathbf{epi}\ g(\ddot{\Pi}(h_t)) = \{(\ddot{\pi}, v) : v \geq g(\ddot{\pi})\}$ onto x_t as a component of $\ddot{\pi}$. Thus the objective $\bar{D}_t(x_t, h_t; \beta)$ in (5) is convex in x_t. ■

Convexity preservation not only facilitates theoretical analysis, but ensures global convergence of numerical methods such as value iteration or policy iteration. If the "adversarial problem" (4) can be solved efficiently, then tractable algorithms may be designed for numerical solutions.

3.1. Competitive Ratio.

Competitive ratio has been applied in many areas, which is preferred when relative regret is more appropriate than absolute regret, and analytical solutions are derived sometimes even with discrete variables (e.g. Wang and Lan 2022). Among the variants of equivalent defitions, the competitive ratio for reward maximization problems can be defined as

$$\gamma^* = \max_{\pi \in \Pi} \min_{\zeta \in \mathcal{U}} r_\zeta(\pi) / r^*(\zeta), \tag{13}$$

where $r^*(\zeta) > 0$ for all $\zeta \in \mathcal{U}$ is assumed in general.

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

Proof: Use backward induction on t. When $t = T + 1$, it is clear that $D_T(h_{T+1}; \beta)$ is continuous in β according to (3), which completes the initial step. The induction step assumes $D_t(h_{t+1}; \beta)$ is continuous in β, then shows the same for $D_{t-1}(h_t; \beta)$. 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 (4). Likewise, $D_{t-1}(h_t; \beta)$ is also continuous with regard to β by (4). ■

The next lemma gives slope bounds for $D(\beta)$, which can establish that $D(\beta)$ strictly increases if $r^*(\zeta) > 0$ for all $\zeta \in \mathcal{U}$.

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

$$r^*(\zeta_{21}^*) \geq \frac{D(\beta_2) - D(\beta_1)}{\beta_2 - \beta_1} \geq r^*(\zeta_{12}^*). \tag{14}$$

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

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

And there is $D(\beta_2) = \max_{\zeta \in \mathcal{U}} \beta_2 r^*(\zeta) - r(\pi_2^*, \zeta) \geq \beta_2 r^*(\zeta_{12}^*) - r(\pi_2^*, \zeta_{12}^*)$. Therefore $D(\beta_2) - D(\beta_1) \geq (\beta_2 - \beta_1)\ r^*(\zeta_{12}^*)$. 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, (14) follows immediately. ■

With continuity and monotonicity from Lemma 1 and 2, it is ready to lay the theoretical foundation for a new approach to competitive ratio analysis.

Theorem 3. *Given $r^*(\zeta) > 0$ for all $\zeta \in \mathcal{U}$ and $r(x, \zeta)$ is bounded below, the equation $D(\beta) = 0$ always has a unique solution $\beta_0 \leq 1$. The competitive ratio $\gamma^* = \beta_0$, and (9) with $\beta = \beta_0$ has the same optimal policies as (13).*

Proof: Positive $r^*(\zeta)$ ensures a strictly increasing $D(\beta)$ by Lemma 2, with $r(x, \zeta)$ bounded below there is $\lim_{\beta \downarrow -\infty} D(\beta) = -\infty$, and $D(1) \geq 0$ is obvious, thus $D(\beta) = 0$ has a unique solution $\beta_0 \leq 1$ by continuity from Lemma 1. Start from (9) and apply Theorem 1:

$$
\begin{aligned}
&\begin{cases} 0 = \min\limits_{\pi \in \Pi} \max\limits_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r(\pi, \zeta) \\ \pi^* \in \operatorname*{argmin}\limits_{\pi \in \Pi} \max\limits_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r(\pi, \zeta) \end{cases} \\
\Leftrightarrow \quad &\begin{cases} 0 = \max\limits_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r(\pi^*, \zeta) \\ \forall \pi \in \Pi : 0 \leq \max\limits_{\zeta \in \mathcal{U}} \beta_0 r^*(\zeta) - r(\pi, \zeta) \end{cases} \\
\Leftrightarrow \quad &\begin{cases} \exists \zeta \in \mathcal{U} : 0 = \beta_0 r^*(\zeta) - r(\pi^*, \zeta) \\ \forall \zeta \in \mathcal{U} : 0 \geq \beta_0 r^*(\zeta) - r(\pi^*, \zeta) \\ \forall \pi \in \Pi, \exists \zeta \in \mathcal{U} : 0 \leq \beta_0 r^*(\zeta) - r(\pi, \zeta) \end{cases} \\
\Leftrightarrow \quad &\begin{cases} \exists \zeta \in \mathcal{U} : \beta_0 = r(\pi^*, \zeta)/r^*(\zeta) \\ \forall \zeta \in \mathcal{U} : \beta_0 \leq r(\pi^*, \zeta)/r^*(\zeta) \\ \forall \pi \in \Pi, \exists \zeta \in \mathcal{U} : \beta_0 \geq r(\pi, \zeta)/r^*(\zeta) \end{cases} \\
\Leftrightarrow \quad &\begin{cases} \beta_0 = \min\limits_{\zeta \in \mathcal{U}} r(\pi^*, \zeta)/r^*(\zeta) \\ \forall \pi \in \Pi : \beta_0 \geq \min\limits_{\zeta \in \mathcal{U}} r(\pi, \zeta)/r^*(\zeta) \end{cases}
\end{aligned}
$$

$$\Leftrightarrow \quad \begin{cases} \beta_0 = \max\limits_{\pi\in\Pi} \min\limits_{\zeta\in\mathcal{U}} r(\pi,\zeta)/r^*(\zeta) = \gamma^* \\ \pi^* \in \operatorname{argmax}\limits_{\pi\in\Pi} \min\limits_{\zeta\in\mathcal{U}} r(\pi,\zeta)/r^*(\zeta) \end{cases}$$

As it can go both ways, the theorem is established. ■

Note that Averbakh (2005) presents some similar results for single-stage problems, which are extended to multistage problems here. So the ARM criterion recovers the relative regret criterion if β is set to the competitive ratio, which could be negative for some problems. The competitive ratio is nonnegative if and only if $D(0) \leq 0$. A new approach to competitive ratio analysis comes straight out of Theorem 3. If an analytical expression for $D(\beta)$ exists, then the competitive ratio can be found by simply solving $D(\beta) = 0$. This approach is generally simpler than directly dealing with the ratio in (13), as illustrated by applying it to one-way trading later.

3.2. Conservatism Control

Recall from earlier discussions that overconservatism can be caused by a criterion obsessed with the worst scenario while ignoring the opportunities in all others. The problem can be exacerbated if the opportunities are ignored in highly likely scenarios, such as a typical scenario ζ^* consisting of most likely values estimated by experts. The ARM criterion may offer a family of robust policies π^*_β by solving (9) for any particular β, so that a choice can be made to best capture opportunities and mitigate overconservatism. When the distribution of ζ is known, a suitable β can be determined by

$$\max_{\beta} E_\zeta \; r(\pi^*_\beta, \zeta), \tag{15}$$

which provides a robust solution with least loss in expected reward. When distribution on scenarios is available but a performance guarantee is highly

desireable, (15) chooses a robust solution with the highest expected reward.

Of course, distributions are unavailable for RO, but (15) lends to heuristics to choose β for conservatism control by exploiting most likely values provided by experts. Suppose $r(x, \zeta)$ is continuous so that scenarios near ζ^* have similar rewards as ζ^*, and $r^*(\zeta^*)$ could also be close to the most likely ex post optimal $\hat{r}^*$. If a policy π has high reward for ζ^*, then it also has high reward for nearby scenarios due to continuity, which is likely to boost the expected reward, thus a β may be found by solving $\max_\beta r(\pi^*_\beta, \zeta^*)$ instead. Surely, a single scenario may not be as representative as a family of scenarios $U(\hat{r}^*)$. For such scenarios, the reward guarantee $\check{r}(\hat{r}^*, \beta)$ may serve as a proxy for the expected reward, and a β may be chosen for an optimal reward guarantee (ORG) $\check{r}^*(r) = \max_\beta \check{r}(r, \beta)$ or

$$\check{r}^*(r) = \max_\beta \beta r - D(\beta), \tag{16}$$

where r may be set to $\hat{r}^*$ or $r^*(\zeta^*)$, and the ORG $\check{r}^*(r)$ is actually the convex conjugate of $D(\beta)$. By this heuristic the ARM criterion may control conservatism to best realize potential opportunities in a family of representative scenarios $U(\hat{r}^*)$. In the case of adapting to DRO with confidence intervals of distribution parameters, $\hat{r}^*$ would be the optimal expected reward given the most likely distribution parameters.

Let $r^*_- = \min_\zeta r^*(\zeta)$ and $r^*_+ = \max_\zeta r^*(\zeta)$ if $r^*(\zeta)$ is bounded, and let $\beta^*_-(r) = \inf B^*(r)$ and $\beta^*_+(r) = \sup B^*(r)$, where $B^*(r) = \{\beta : \beta r - D(\beta) = \check{r}^*(r)\}$. When $r \in \{r^*_-, r^*_+\}$, an upper bound for $\check{r}(r, \beta)$ is

$$\bar{r}(r) = \max_{\pi \in \Pi} \min_{\zeta \in U(r)} r(\pi, \zeta), \tag{17}$$

which comes from

$$\begin{aligned} D(\beta) &= \min_{\pi\in\Pi}\max_{\zeta\in\mathcal{U}} \beta r^*(\zeta) - r(\pi,\zeta) \\ &\geq \min_{\pi\in\Pi}\max_{\zeta\in U(r)} \beta r^*(\zeta) - r(\pi,\zeta) \\ &= \beta r - \bar{r}(r). \end{aligned}$$

Theorem 4. *If $r(x,\zeta)$ is bounded, the ORG $\check{r}^*(r)$ and the maximizers $\beta_-^*(r)$ and $\beta_+^*(r)$ in (16) have these properties:*

i. For $r_1 > r_0$, there is $\beta_-^*(r_1) \geq \beta_+^*(r_0)$, the ORG $\check{r}^*(r)$ is convex with $\check{r}^*(r_0) < \check{r}^*(r_1)$ if $\beta_+^*(r_0) > 0$ and $\check{r}^*(r_0) > \check{r}^*(r_1)$ if $\beta_-^*(r_1) < 0$, and

$$\begin{cases} \beta_+^*(r) = -\infty,\ \check{r}^*(r) = +\infty & \text{if } r < r_-^* \\ \beta_-^*(r) = -\infty,\ \check{r}^*(r) \leq \bar{r}(r_-^*) & \text{if } r = r_-^* \\ \beta_-^*(r) > -\infty,\ \beta_+^*(r) < +\infty & \text{if } r \in (r_-^*, r_+^*) \\ \beta_+^*(r) = +\infty,\ \check{r}^*(r) \leq \bar{r}(r_+^*) & \text{if } r = r_+^* \\ \beta_-^*(r) = +\infty,\ \check{r}^*(r) = +\infty & \text{if } r > r_+^* \end{cases}$$

ii. For $r \in [r_-^*, r_+^*]$, the absolute guarantee gap (AGG) of $G(r) = r - \check{r}^*(r) \geq 0$, is concave and strictly increases (decreases) in r if $\beta_+^*(r) < 1$ (if $\beta_-^*(r) > 1$). When $r \geq r_-^* > 0$, the relative guarantee gap (RGG) of $G(r)/r$ strictly increases (decreases) in r if $\beta_+^*(r) < \gamma^*$ (if $\beta_-^*(r) > \gamma^*$).

Proof: These properties are proved as follows. i. Let $\beta_i^* \in B^*(r_i)$ for

$i = 0, 1$. Assume $\beta_1^* < \beta_0^*$, optimality of β_0^* gives $\check{r}(r_0, \beta_1^*) \leq \check{r}(r_0, \beta_0^*)$

$$\begin{aligned}
\Rightarrow \quad & r_0\beta_1^* - D(\beta_1^*) && \leq r_0\beta_0^* - D(\beta_0^*) \\
\Rightarrow \quad & D(\beta_0^*) - D(\beta_1^*) && \leq r_0(\beta_0^* - \beta_1^*) \\
\Rightarrow \quad & D(\beta_0^*) - D(\beta_1^*) && < r_1(\beta_0^* - \beta_1^*) \\
\Rightarrow \quad & r_1\beta_1^* - D(\beta_1^*) && < r_1\beta_0^* - D(\beta_0^*),
\end{aligned}$$

leading to $\check{r}(r_1, \beta_1^*) < \check{r}(r_1, \beta_0^*)$, a contradiction to the optimality of β_1^*, which proves $\beta_1^* \geq \beta_0^*$, implying $\beta_-^*(r_1) \geq \beta_+^*(r_0)$ when $\beta_1^* = \beta_-^*(r_1), \beta_0^* = \beta_+^*(r_0)$.

Note that $\check{r}^*(r)$ is convex conjugate of $D(\beta)$, a pointwise maximum of a family of strictly increasing (decreasing) affine functions with slope $\beta_+^*(r_0) > 0$ ($\beta_-^*(r) < 0$) of r, hence convex and strictly increasing after r_0 (decreasing before r_1) with monotonicity of $\beta_-^*(r)$ and $\beta_+^*(r)$.

By Lemma 2, there is $r_-^* \leq (D(\beta_1) - D(\beta_0))/(\beta_1 - \beta_0) \leq r_+^*$ for $\beta_1 > \beta_0$, which gives $D(\beta_0) \leq D(\beta_1) - (\beta_1 - \beta_0)r_-^*$, so that $\check{r}(r, \beta_0) \geq \check{r}(r, \beta_1) + (\beta_1 - \beta_0)(r_-^* - r)$. Clearly, if $r < r_-^*$, there is $\lim_{\beta_0 \downarrow -\infty} \check{r}(r, \beta_0) = \infty$, giving $\check{r}^*(r) = \infty$ and $\beta_+^*(r) = -\infty$. If $r = r_-^*$ then $\check{r}(r, \beta)$ increases as $\beta \downarrow -\infty$, thus $\beta_-^*(r) = -\infty$ and $\check{r}^*(r_-^*) = \lim_{\beta \downarrow -\infty} \beta r_-^* - D(\beta)$, which exists as it increases with an upper bound $\bar{r}(r_-^*)$. For the case of $r \in (r_-^*, r_+^*)$, let π_β^* be the optimal policy, ζ_β^* the worst scenario, and $\zeta_-^* \in U(r_-^*)$. Clearly,

$$D(\beta) = \beta r^*(\zeta_\beta^*) - r(\pi_\beta^*, \zeta_\beta^*) \geq \beta r^*(\zeta_-^*) - r(\pi_\beta^*, \zeta_-^*) \geq (\beta - 1)r_-^*.$$

Then $r(r, \beta) \leq \beta(r - r_-^*) + r_-^*$, which goes to $-\infty$ as $\beta \downarrow -\infty$, thus $\beta_-^*(r) > -\infty$. The results involving r_+^* follow similarly.

ii. Rewrite $G(r)$ into $G(r) = \min_{\beta \geq 0} D(\beta) + (1 - \beta)r$, a pointwise minimum of a family of strictly increasing (decreasing) affine functions in r

when $\beta_+^*(r) < 1$ ($\beta_+^*(r) > 1$). Rewrite $G(r)/r$ for $r > 0$ into $G(r)/r = \min_{\beta \geq 0} D(\beta)/r - \beta + 1$, a pointwise minimum of a family of increasing (decreasing) functions in r when $D(\beta) < 0$ ($D(\beta) > 0$), as $D(\beta)$ strictly increases with $r_-^* > 0$ by Lemma 2. By Theorem 3, when $\beta_+^*(r) < \gamma^*$ ($\beta_-^*(r) > \gamma^*$), there is $D(\beta_+^*(r)) < 0$ ($D(\beta_-^*(r)) > 0$), and $G(r)/r$ should strictly increase (decrease) in the neighborhood of r. ■

Theorem 4 reveals that the ARM criterion best captures opportunities around a bigger $\hat{r}^*$ by a bigger β^*, demonstrating the conservatism moderating role of β. In practical applications, there is $r = \hat{r}^* \in [r_-^*, r_+^*]$ and thus $\check{r}^*(r)$ is bounded, yet the direction of its changes depends on the sign of $\beta_-^*(r)$ and $\beta_+^*(r)$. The AGG and RGG are indicators of efficiency to capture opportunities, where RGG is appropriate for considering relative losses. High efficiency (small AGG or RGG) when r is near the lower or higher end of $[r_-^*, r_+^*]$ may be interpreted as opportunities are easier to capture when they are cornered to either ends. Theorem 4 requires almost nothing on $D(\beta)$, but if $D(\beta)$ is continuous and convex, then $D(\beta)$ is also the convex conjugate of $\check{r}^*(r)$. Note that $\beta r^*(\zeta) - r(\pi, \zeta)$ is linear in β, thus $F(\beta; \pi) = \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r(\pi, \zeta)$ is convex in β for a given policy π. But $D(\beta) = \min_{\pi \in \Pi} F(\beta; \pi)$ is not necessarily convex in β, and conditions are required to make it convex.

Theorem 5. *With consistent convexity, the optimal regret guarantee $D(\beta)$ is convex in β, and $D(\beta)$ is strictly convex if $r(x, \zeta)$ is strictly concave in x for all $\zeta \in \mathcal{U}$.*

Proof: Let $\ddot{\pi}_i^* \in \ddot{\Pi}$ be an optimal pruned policy for $\beta_i, i = 1, 2$, and let $\ddot{\pi}' = \lambda \ddot{\pi}_1 + (1 - \lambda)\ddot{\pi}_2$ for any $\lambda \in (0, 1)$. By the concavity of $r(x, \zeta)$, there is

$r(\ddot{\pi}', \zeta) \geq \lambda r(\ddot{\pi}_1^*, \zeta) + (1-\lambda) r(\ddot{\pi}_2^*, \zeta)$ for all $\zeta \in \mathcal{U}$. Let $\beta = \lambda\beta_1 + (1-\lambda)\beta_2$ and proceed as follows:

$$\begin{aligned}
D(\beta) &= \min_{\ddot{\pi} \in \ddot{\Pi}} \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r(\ddot{\pi}, \zeta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - r(\ddot{\pi}', \zeta) \\
&\leq \max_{\zeta \in \mathcal{U}} \beta r^*(\zeta) - (\lambda r(\ddot{\pi}_1^*, \zeta) + (1-\lambda) r(\ddot{\pi}_2^*, \zeta)) \\
&\leq \lambda \left(\max_{\zeta \in \mathcal{U}} \beta_1 r^*(\zeta) - r(\ddot{\pi}_1^*, \zeta) \right) + \\
&\quad (1-\lambda) \left(\max_{\zeta \in \mathcal{U}} \beta_2 r^*(\zeta) - r(\ddot{\pi}_2^*, \zeta) \right) \\
&= \lambda D(\beta_1) + (1-\lambda) D(\beta_2).
\end{aligned}$$

Similarly, strict concavity comes by $r(\ddot{\pi}', \zeta) > \lambda r(\ddot{\pi}_1^*, \zeta) + (1-\lambda) r(\ddot{\pi}_2^*, \zeta)$. ■

If $D(\beta)$ is strictly convex and differentiable with a continuous and strictly increasing $D'(\beta)$, then $\beta^*(r)$ for (16) is simply the inverse of $D'(\beta)$ from the first order condition

$$\frac{\partial \check{r}(r, \beta)}{\partial \beta} = r - D'(\beta) = 0. \tag{18}$$

So the ARM criterion can indeed adjust the level of conservatism to capture potential opportunies, while maintaining all major advantages of RO. This mechanism could be highly valuable as less conservative solutions are recommended, more performance can be expected, while only requiring most likely values as additional information from experts. According to Vinod (2021), for example, overconservatism is the main issue preventing the adoption of robust revenue management in airlines, as the profit margin for airlines is razor-thin, and even a 1% change in revenue could make a huge difference.

4. One-way Trading.

In this section the ARM criterion is applied to the one-way trading problem to demonstrate its properties and potential, such as the new approach to competitive ratio analysis and the effectiveness of the conservatism control hueristic. The one-way trading problem has been richly studied with both competitive ratio (El-Yaniv et al. 2001) and absolute regret (Wang et al. 2016), which ideally serve as targets of comparison. Closed-form analytical solutions are derived with the ARM criterion, yielding the result of Wang et al. (2016) as a special case with $\beta = 1$ while the derivation process is not much more complex than theirs. The competitive ratio is directly found via the new approach, in contrast to the lengthy and complex derivation process in El-Yaniv et al. (2001) that heavily depends on acute intuition and deep insights that call for great talents. Finally, numerical simulations are conducted to verify that the ARM criterion can indeed offer smooth control of conservatism, and the heuristic can determine an appropriate level of conservatism for improved performances.

4.1. Problem Formulation.

Consider selling a certain amount of divisible goods (such as gasoline) in periods $t = 1, \cdots, T$, while the price fluctuates in the range of $[m, M]$. A single price $p_t \in [m, M]$ is first revealed in each period, then as a price-taker the trader sells at p_t an amount $x_t \geq 0$ out of the remaining stock, without knowing any future prices. In the last period T, the trader must sell out whatever remains. The goal is to maximize the total sales revenue.

It is a multistage problem, in which the stages naturally coincide with

periods. A scenario ζ 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 to sell is 1 unit, and the action is $x = (x_1, \cdots, x_T)$ with $X = \{x \geq 0 : \sum_{t=1}^{T} x_t = 1\}$. 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 stock to sell given h_t, but in the last period $X_T(h_T) = [q_T, q_T]$. The reward is accrued over time, so let $r_t = \sum_{s=1}^{t-1} p_s x_s$ for the rewards accrued over h_t, and 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. In the end (3) becomes

$$D_T(h_{T+1}; \beta) = \beta \hat{p}_{T+1} - r_{T+1}. \tag{19}$$

By tradition, in a stage t the uncertain price p_t is revealed first, then an action x_t is taken, which differs from the standardized formulation in (5):

$$D_{t-1}(h_t; \beta) \quad = \quad \max_{p_t \in [m,M]} \min_{x_t \in X_t(h_t)} D_t(h_{t+1}; \beta). \tag{20}$$

As a dummy decision can be added to have it standardized, this difference is superficial, and all results in Section 3 remain valid.

4.2. Analytic Solution.

The analysis starts from the last period T and works backwards. It is sufficient to consider $\beta \geq 0$ for comparison with related work, while keeping things simple. In the last period there is $x_T = q_T$, and (20) 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 $p_T = M$, thus

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

where $R_t = r_t + mq_t$ for $t = 1, \cdots, T$ is the lower bound on r_{T+1} given h_t, and $P_1(y)$ is an auxiliary quantity-to-price function defined as

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

with $y^{+j} = \max^j(0, y)$ for the positive part of y raised to the j^{th} power. Let $P_j^-(y) = q$ be the inverse of $y = P_j(q)$ for $q \in [0, \beta j]$. The trivial case of $\beta = 0$ is found by the limit as $\beta \downarrow 0$. Continue on with (20) for $t = T-1, \cdots, 1$ by backward induction, analytical solutions can be obtained.

Theorem 6. *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*

$$D_{t-1}(h_t;\beta) = \beta \max(\hat{p}_t, P_{1+T-t}(q_t)) - R_t, \tag{21}$$

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

$$q_{t+1}^* = \min(q_t, P_{T-t}^-(\hat{p}_{t+1})),\ t = 1, \cdots, T-1. \tag{22}$$

Proof: By backward induction. For the initial step with $t = T$, it is easily verified. For the induction step, assume (21) holds in period $t+1 \leq T$ with

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

and prove it also holds in period t. For the minimization nested in (20), let

$$\begin{aligned}\bar{D}_t(h_t, p_t; \beta) &= \min_{x_t \in X_t(h_t)} D_t(h_{t+1}; \beta) \\ &= \min_{q_{t+1} \in [0, q_t]} \beta \max(\hat{p}_{t+1}, P_n(q_{t+1})) - R_{t+1}, \end{aligned} \tag{23}$$

with $n = T - t$, $q_{t+1} = q_t - x_t$, and $R_{t+1} = r_{t+1} + mq_{t+1}$. First note that

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

which means $P_n(q)$ is monotone and there is $P_n(q_{t+1}) \geq \hat{p}_{t+1}$ if $q_{t+1} \leq P_n^-(\hat{p}_{t+1})$ and $P_n(q_{t+1}) \leq \hat{p}_{t+1}$ otherwise. Therefore,

$$D_t(h_{t+1}; \beta) = \begin{cases} \beta P_n(q_{t+1}) - R_{t+1} & q_{t+1} \leq P_n^-(\hat{p}_{t+1}) \\ \beta \hat{p}_{t+1} - R_{t+1} & q_{t+1} > P_n^-(\hat{p}_{t+1}) \end{cases} \tag{24}$$

$$\frac{\partial D_t(h_{t+1}; \beta)}{\partial q_{t+1}} = \begin{cases} p_t - m + \beta P_n'(q_{t+1}) & q_{t+1} < P_n^-(\hat{p}_{t+1}) \\ p_t - m & q_{t+1} > P_n^-(\hat{p}_{t+1}) \end{cases} \tag{25}$$

Note that 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 with $q_{t+1} > P_n^-(\hat{p}_{t+1})$, there is $p_t - m \geq 0$. It is clear that (22) is an optimal solution to (23), which from (24) gives

$$\bar{D}_t(h_t, p_t; \beta) = \beta P_n(q_{t+1}^*) - (r_{t+1} + mq_{t+1}^*). \tag{26}$$

Let $\bar{p}_t = \max(\hat{p}_t, P_n(q_t)) \in [m, M]$, and from (20) there is

$$\begin{aligned} D_{t-1}(h_t; \beta) &= \max_{p_t \in [m, M]} \bar{D}_t(h_t, p_t; \beta) \\ &= \max \begin{pmatrix} \max_{p_t \in [m, \bar{p}_t]} \bar{D}_t(h_t, p_t; \beta) \\ \max_{p_t \in [\bar{p}_t, M]} \bar{D}_t(h_t, p_t; \beta) \end{pmatrix} \end{aligned} \tag{27}$$

For the branch with $p_t \in [m, \bar{p}_t]$ in (27), 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 (22) simplifies to $q_{t+1}^* = P_n^-(\hat{p}_{t+1})$, 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 (22) simplifies to $q_{t+1}^* = q_t$, thus $P_n(q_{t+1}^*) = P_n(q_t) = \bar{p}_t$. So there is $P_n(q_{t+1}^*) = \bar{p}_t$ in both cases, and (26) becomes $\bar{D}_t(h_t, p_t; \beta) = \beta\bar{p}_t - (r_{t+1} + mq_{t+1}^*) = \beta\bar{p}_t - r_t - p_t x_t^* - mq_{t+1}^*$, 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 with $p_t \in [\bar{p}_t, M]$ in (27), 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 (22) simplifies to $q_{t+1}^* = P_n^-(\hat{p}_{t+1})$. Therefore $P_n(q_{t+1}^*) = \hat{p}_{t+1} = p_t$, and (26) simplifies to $\bar{D}_t(h_t, p_t; \beta) = \beta p_t - r_t - p_t x_t^* - mq_{t+1}^* = \beta p_t - r_t - p_t(q_t - q_{t+1}^*) - mq_{t+1}^* = (\beta - q_t + q_{t+1}^*)p_t - mq_{t+1}^* - r_t = (\beta - q_t + q_{t+1}^*)P_n(q_{t+1}^*) - mq_{t+1}^* - r_t = d(P_n^-(p_t))$, where $d(z) = (\beta - q_t + z)P_n(z) - mz - r_t, z \in [0, 1]$, with a derivative $d'(z) = (\beta - q_t + z)P_n'(z) + P_n(z) - m$. 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 $z^* = nq_t/(n+1) < q_t$ solves $\max_{z \in [0,1]} d(z)$, which gives

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

Consider two cases with $\bar{D}_t(h_t, p_t; \beta) = d(P_n^-(p_t))$ for $p_t \in [\bar{p}_t, M]$. 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 (27) there is

$$D_{t-1}(h_t; \beta) = \max(\beta\bar{p}_t - R_t, d(z^*)). \tag{28}$$

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

$$\begin{aligned}\max_{p_t \in [\bar{p}_t, M]} \bar{D}_t(h_t, p_t; \beta) &= \max_{p_t \in [\bar{p}_t, M]} d(P_n^-(p_t)) \\ &= \max_{z \in [0, P_n^-(\bar{p}_t)]} d(z) \\ &\leq \max_{z \in [0, q_t]} d(z) = d(z^*).\end{aligned}$$

As $P_n(z^*) \geq P_{n+1}(q_t)$, there is $\bar{p}_t \geq P_{n+1}(q_t)$. So $d(z^*) = \beta P_{n+1}(q_t) - R_t \leq \beta \bar{p}_t - R_t$, and by (27) there is $D_{t-1}(h_t; \beta) = \beta \bar{p}_t - R_t$, and (28) remains valid. Therefore, in both cases proceed from (28) and take note of $\bar{p}_t = \max(\hat{p}_t, P_n(q_t))$ and $P_n(q_t) \leq P_{n+1}(q_t)$:

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

As $n = T - t$, clearly (21) 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:*

$$D(\beta) = \beta(M - m)\left(1 - \frac{1}{\beta T}\right)^{+T} - (1 - \beta)m, \tag{29}$$

Proof: In the first period, there is $q_1 = 1, r_1 = 0, \hat{p}_1 = m$. Use these in (21) and simplify to have the result. The convexity of $D(\beta)$ is a consequence of the consistent convexity of the one-way trading problem and Theorem 5. ■

The result of Wang et al. (2016) is a special case of Theorem 6 with $\beta = 1$, and the proof for this more general result requires more general treatments.

Theorem 6 easily leads to a tremendously simplified derivation of the competitive ratio, as compared to the truly ingenious and highly complicated analysis of El-Yaniv et al. (2001).

Corollary 2. *The competitive ratio defined in (13) for the one-way trading problem is the unique root* β_0 *of* $D(\beta) = 0$ *as defined in (29).*

Proof: As $r^*(\zeta) \geq m > 0$, it follows from Theorem 3. ■

This result agrees perfectly with El-Yaniv et al. (2001), except that they define competitive ratio as its inverse. Their analysis is much more involved and heavily relies on insights of the worst case price paths, which can be deduced straightforwardly once β_0 is known, in a way similar to what is done in Wang et al. (2016).

Corollary 3. *As* β *increases, the optimal trading policy gets more optimistic and aggressive: Other things being equal, it takes on more risks by trading less now and reserving more for the future, which means* $\forall h_t \in H_t, \forall p_t \in [m, M]$ *there is*

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

Proof: Consider the quantity reserved for future q_{t+1}^* in (22) 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 displays theoretically the continuous moderation of conservatism by β as the optimal policy gets more optimistic and aggressive for bigger β values.

4.3. Numerical Study.

The effects of heuristics and the fine control of conservatism is further studied numerically on the one-way trading problem. Here is the basic setup. The instance has $T = 5$ periods, and the prices are in $[m, M] = [1, 3]$. The prices for all periods are independent and identically distributed (IID) with a Beta(a, b) distribution on $[m, M]$ with shape parameters $a = 3.5, b = 1.5$. Suppose experts accurately estimate the most likely value for ex post optimal reward $\hat{r}^* = 2.897$, by which the best $\beta^* = 2.58$ is found numerically by (18). Optimal policies for $\beta \in \{i/100 : i = 0, \cdots, 400\}$ are computed from (22), and all policies are executed on the same sequence of randomly generated prices to obtain the overall rewards. The whole process is repeated $N = 10,000$ times and the average and standard deviation of the reward for each β is calculated, by which a 99% confidence interval (CI) for the average reward is computed.

Fig. 1 has the results. Firstly, the heuristic works quite well, with the guarantee $\check{r}^*(\beta) = \beta\hat{r}^* - D(\beta)$ being a fairly good proxy for the average reward to find the maximizing β. The empirical maximizer $\beta = 0.259$ is found on the average reward curve in the spirit of (15), while the heuristic finds $\beta = 2.57$ as the maximizer on the best guarantee curve, which are very close to each other with almost identical average reward. These policies are entered in Table 1 as "heuristic" and "empirical" policies with other special ones: the maximin policy with $\beta = 0$, the relative regret policy with $\beta = \gamma^* = 0.72$, and the absolute regret policy with $\beta = 1$, while the second last row has the policy that maximizes expected reward (interested readers are referred to Appendix B of Wang and Lan (2022) for more details)

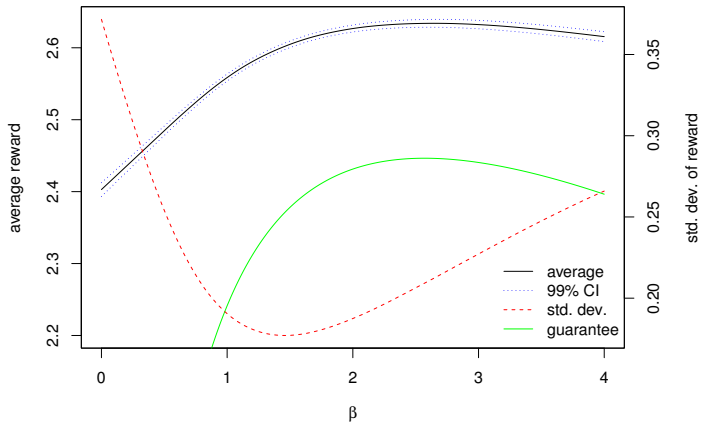

Figure 1: Average reward, guarantee, and standard deviation of reward. The average reward is unimodal, peaks at $\beta = 2.59$ with a value of 2.636. The guarantee for $\hat{r}^*$ peaks at $\beta = 2.57$, very close to 2.59. The standard deviation is unimodel with a valley at $\beta = 1.45$, achieving a minimal value of 0.177.

and the last row has the ex post optimal policy. Benchmarking against the expected reward maximizing policy, the absolute and relative gaps are listed in the table, where the heuristic has a gap of 3.2% and improves the average reward by roughly 9% from maximin, 4% from relative regret, and 3% from absolute regret. Such improvements can make a big difference in some practical applications, such as airline revenue management. Note that the heuristic makes more significant improvement over the other policies if the heuristic β is further away from theirs, which helps choose the values of $a = 3.5, b = 1.5$ for Beta(a,b) to show off the potential of the ARM criterion and the heuristic.

It seems in this experiment that the heuristic finds a sweet spot in between extreme conservatism and aggressiveness. The extremely aggressive case of

Policy (β)	Average±99% CI	Gap	Gap%
maximin (0.00)	2.397±0.010	0.327	12.0%
relative (0.72)	2.519±0.006	0.205	7.5%
absolute (1.00)	2.560±0.005	0.165	6.0%
heuristic (2.57)	2.636±0.005	0.088	3.2%
empirical (2.59)	2.636±0.005	0.088	3.2%
max expected	2.725±0.006	—	—
ex post optimal	2.790±0.004	-0.066	-2.4%

Table 1: Benchmark the ARM policies of various β values.

$\beta \uparrow \infty$ (absent in Fig. 1) almost only sells in the last period by (22), while the extremely conservative case of $\beta = 0$ almost only sells in the first period, giving them identical average reward and standard deviation. Thus both extremes give poor performance with low expected reward and high overall risk. By adjusting the β value of the ARM criterion, the level of conservatism can be fine-tuned to match the situation for higher rewards and lower risks.

Next, the heuristic is observed in a broader perspective for its capacity of fine control of conservatism. The same setup is used except for different shape parameters in Beta(a,b) for $a \in [0.1, 3.9]$ with a step size of 0.1 and $b = 5 - a$: the range avoids $b \leq 1$ because the density would diverge at M, causing $\beta^* = \infty$ for the heuristic. As bigger a value is condusive to the probability for higher prices and more chances of good opportunities, a bigger β should be employed for a less conservative robust policy in theory.

Fig. 2 has the average rewards and corresponding β for the heuristic and empirical policy. The heuristic policy gives an expected reward very close to

that of the empirical policy in general, despite the fairly large discrepancy in β at $a = 3.9$. The β for both the heuristic and empirical policy indeed steadily increases as a gets bigger, illustrating continuous control of conservatism by β to best catch increasingly better opportunities, as predicted in theory. The average reward of the heuristic does start to fall behind a little as $a \uparrow 4$ (see the round head of heuristic reward curve in the top right corner with a reward of 2.75 at $a = 3.9$, dropping 1% from the empirical), while the heuristic β significantly overshoots the empirical β rapidly. It indicates that the mode $\hat{r}^*$ is getting less representative, as the Beta distribution approaches the point of divergence at M with $a = 4, b = 1$, resulting in $\hat{r}^* = M$ and $\beta^* = \infty$, according to Theorem 4 with $[r^*_-, r^*_+] = [m, M]$.

A quick fix for reduced average reward due to impaired representativeness of $\hat{r}^*$ near the point of divergence is to expand the set of representative scenarios to $U(r, \delta) = \{\zeta : r^*(\zeta) \in [r-\delta, r+\delta] \cap [r^*_-, r^*_+]\}$ to determine a right β, where $r = \hat{r}^*, \delta = (r^*_+ - r^*_-) \cdot \eta$, and $\eta = 5\%$ in the experiment. A simple method to determine β is employed: simply use (18) with the middle point of the defining interval for $U(r, \delta)$: $r_m(r, \delta) = (\max(r^*_-, r-\delta) + \min(r^*_+, r+\delta))/2$. In Fig. 2 this method is labeled "midpoint", whose average reward sticks with that of the empirical throughout, and their β values are always close to each other.

5. Conclusion.

The ARM criterion proposed in this paper provides fine control of conservatism by the CCP (β), while maintaining all the major advantages of RO. It minimizes the β-adjusted regret guarantee, from which a reward guarantee

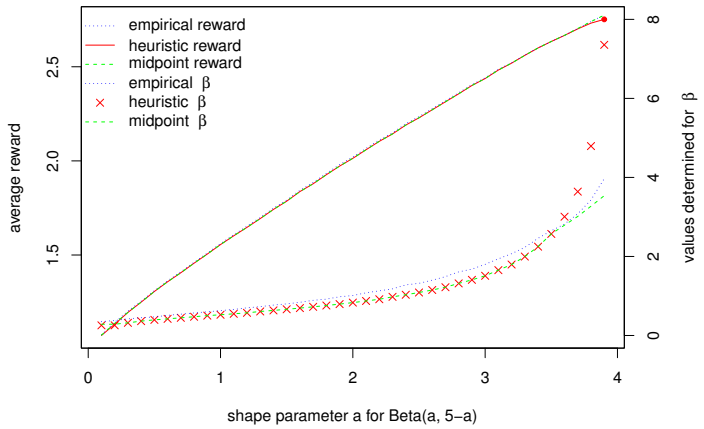

Figure 2: The heuristic and empirical policy under a continous shape shifting scheme for the Beta distribution to simulate environments with increasingly better opportunities.

for scenarios can be derived. And convexity is preserved even for multistage problems, a property great for computational tractability and theoretical analysis. Distributions on uncertainty is not required, only the most likely values are needed from experts to calibrate β heuristically for the right level of conservatism to best catch opportunities and improve the performance. The heuristic is based on the mechanism that as β increases, the ARM criterion will recommend solutions with better reward guarantees for scenarios bearing more opportunities. It is possible to adapt the ARM criterion and the heuristic to DRO as well. 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. These theoretical results also lead to a new approach for competitive ratio analysis, which may be much simpler

than the traditional approach, as is observed in the analysis of the one-way trading problem.

The ARM criterion is applied to the robust one-way trading problem to demonstrate its potential. Closed-form solution is obtained, from which the competitive ratios is quickly derived by the new approach. Analysis of the closed-form solution shows that the optimal policy gets more aggressive as β increases. Numerical experiments on one-way trading are designed to illustrate fine control of conservatism, with significant benefits of the heuristic.

This study of the ARM criterion only serves as a starting point for future research. First of all, applying it to pratical problems in various areas is the thrust for further theoretical development. Conceivably, innovative methods may be developed to find an appropriate β in practice. Note that it can also be applied when there is rich historical data to estimate distributions, but a performance guarantee is highly desired. Researches on linear problems with the ARM criterion can be fruitful, as progresses are made in this regard on the absolute and relative regret criterion by Poursoltani and Delage (2021). Finally, it can be practically and theoretically fruitful to apply the ARM criterion with DRO.

References

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

Bellman, R. (1954). The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515.

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

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. 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.

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.

Hurwicz, L. (1951). Optimality criteria for decision making under ignorance. Technical report, Cowles Commission discussion paper, statistics.

Kalaı, R., Lamboray, C., and Vanderpooten, D. (2012). Lexicographic α-robustness: An alternative to min–max criteria. *European Journal of Operational Research*, 220(3):722–728.

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

Kuhn, D., Shafiee, S., and Wiesemann, W. (2025). Distributionally robust optimization. *Acta Numerica*, 34:579–804.

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

Murty, K. G. and Kabadi, S. N. (1987). Some np-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129.

Nesterov, Y. and Nemirovskii, A. (1994). *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics.

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.

Savage, L. J. (1954). *The foundations of statistics*. Courier Corporation.

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

Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE transactions*, 38(7):547–564.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5):1154–1157.

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.

Wald, A. (1950). *Statistical decision functions*. Wiley publications in statistics. New York, Wiley.

Wang, W. and Lan, Y. (2022). Robust one-way trading with limited number of transactions and heuristics for fixed transaction costs. *International Journal of Production Economics*, 247:108437.

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.

Zhen, J., Kuhn, D., and Wiesemann, W. (2025). A unified theory of robust and distributionally robust optimization via the primal-worst-equals-dual-best principle. *Operations Research*, 73(2):862–878.