Adjustable Regret for Continuous Control of Conservatism and Competitive Ratio Analysis

Yingjie Lan
Peking University, ylan@pku.edu.cn,

A major issue of the increasingly popular robust optimization is the tendency to produce overly conservative solutions. This paper proposes a new parameterized robust criterion to offer smooth control of conservatism without tampering with the uncertainty set. Unlike many other intractable criteria, its tractability is attained for common types of linear problems by reformulating them into traditional linear robust optimization problems. Many properties of it are studied to help analyze multistage robust optimization problems for closed-form solutions and give rise to a new approach to competitive ratio analysis. Finally, the new criterion is applied to the well-studied robust one-way trading problem to demonstrate its potential. A closed-form solution is obtained, which not only facilitates a numerical study of its smooth control of conservatism, but leads to a much simpler competitive ratio analysis.

Key words: robust optimization; decision criteria; over-conservatism; tractability; competitive ratio analysis; one-way trading

1. Introduction. Robust optimization (RO) in a broad sense deals with decision-making under uncertainty without knowing an exact distribution, which is, in contrast, a prerequisite for stochastic programming. RO is often considered more practical and powerful in many real applications where an exact distribution is difficult to correctly estimate, due to little or inaccurate data, or nonstationarity of the underlying stochastic process. A survey of RO in the broad sense is given by Gabrel et al. (2014), which covers topics from distributionally RO to the traditional or narrow-sensed RO. The traditional RO traces back to Soyster (1973), which proposes a robust counterpart for a linear programming (LP) model to address the concern on feasibility as constraints may be violated when the data vary. A large branch of studies follows Soyster (1973) to derive tractable reformulations, providing insights into robust solutions as well as probabilistic guarantees of constraint violation (see Ben-Tal and Nemirovski 2008 and Bertsimas et al. 2011 for comprehensive surveys and Ben-Tal et al. 2009 for a book treatment on this topic).
Equally important is the concern on optimality, which raises the issue of over-conservatism, noted very early in the development of RO, with much effort made to mitigate it. The robust counterpart of Soyster (1973) is found to produce solutions that are too conservative, sacrificing too much performance to ensure feasibility for all scenarios considered. A justification for such complete protection from all scenarios may be made in applications where infeasibility can literally cause disasters like a doomed satellite launch or a destroyed unmanned robot. But it is less advisable in business applications where adverse events like low demand or supply do not bring about such disastrous consequences. In the latter case, the protection can be made instead against a smaller and more flexible uncertainty set with a lower probabilistic guarantee of feasibility. Significant progress in the theory of RO was then made independently in this regard for models with ellipsoidal uncertainties by Ben-Tal and Nemirovski (1998, 1999, 2000), El-Ghaoui and Lebret (1997), and El Ghaoui et al. (1998). A practical drawback of such an approach is that it leads to nonlinear models that are more demanding computationally than the linear models in Soyster (1973). To maintain comparable tractability, Bertsimas and Sim (2004) proposes the uncertainty budget that offers full control on the degree of conservatism for every constraint. All these approaches carefully choose and control the uncertainty set to strike a balance between robustness and performance, while trying their best to maintain tractability.

A different strategy to tackle over-conservatism is to utilize criteria well posed to take advantage of favorable scenarios and gain performance as opportunity allows. The traditional RO formulations adopted the commonly used maximin reward (or minimax cost) criterion that optimizes the worst-case objective. However, this criterion tends to recommend overly conservative solutions, which is considered by some as the Achilles’ heel of the traditional RO, and people started to look for less conservative criteria. A well-known alternative is the absolute regret criterion, which minimizes the regret experienced by decision-makers once they realize what would have been the best actions in hindsight. It is proposed by Savage (1951), and axiomatized in Milnor (1951) and Stove (2011). The regret may also be considered in its relative form with the absolute regret in ratio to the optimal objective in hindsight (Kouvelis and Yu 2013), and the resultant relative regret criterion is equivalent to the so-called “competitive ratio,” a popular measure for online optimization (Borodin and El-Yaniv 2005). These three criteria have received most interest in researches and applications as they are easier to analyze and more computationally tractable than other criteria.

These criteria can offer different levels of conservativism, usually in a consistent order. Many studies describe both regret criteria as leading to less “conservative” decisions than those with 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). Note that it does not require any extra knowledge on the uncertainty to achieve this, it is rather due to the characteristics of the criteria.
The maximin reward criterion entirely focuses on the worst-case profit and completely ignores all plausible opportunities for higher profits, which frequently leads to over-conservatism in scenarios with large regret. On the other hand, both absolute and relative regret criteria are better posed to seize such opportunities and deliver a less conservative performance, whilst sacrificing some assurance of worst-case profit. It is further observed by [Lan et al. (2008) and Wang et al. (2016)] that for profit maximization, absolute regret is less conservative than relative regret. This observation corroborates with the findings in the numerical experiments of [Poursoltani and Delage (2021)], which also confirms maximin reward to be the most conservative. Note that only three distinct levels of conservatism are offered by these criteria. To offer a *continuum of choices for smooth control of conservatism*, this paper proposes a new criterion of adjustable regret minimization (ARM) by interpolating between and extrapolating beyond them.

Computational tractability is very important in applications of RO. Although traditional RO formulations with the maximin reward criterion are polynomially solvable in the case of an LP model with box uncertainty for objective coefficients, [Averbakh and Lebedev (2005)] showed that solving the worst-case regret minimization form is strongly NP-hard. Since then extensive efforts are made to develop exact and approximate solution schemes. The most recent development in [Poursoltani and Delage (2021)] reformulates two-stage regret minimization problems into two-stage traditional RO problems, so as to benefit from the tractable solution schemes and theoretical results developed so far for this class of problems (see [Yanıkoğlu et al. 2019] for a recent survey). Parallel results are obtained for ARM, with similar tractability as the absolute regret criterion.

Sometimes closed-form solutions can be derived for regret minimization problems, making it easy to analyze the structural properties and numerically compute the solutions. [Yue et al. (2006) and Perakis and Roels (2008)] give analytical solutions for two-stage newsvendor problems with only one item, absolute regret, and distribution ambiguity. [Lan et al. (2008)] provide closed-form solutions to both absolute and relative regret models for single-leg revenue management problems. [Wang et al. (2016) and Wang and Lan (2022)] analytically solve robust multistage one-way trading problems. To facilitate finding such analytical solutions, the theoretical properties of ARM are analyzed in a multistage setting, with single- and two-stage problems covered as special cases.

The theoretical results also give rise to a new approach to competitive ratio analysis. Usually, competitive ratio analysis considers the ratio directly, which can be much more involved than the absolute regret analysis dealing with the difference as regret. For example, [Lan et al. (2008)] studies both competitive ratio and absolute regret on the same problem setting, where the latter is much easier to analyze than the former. Similar observations can be made by comparing the analyses of one-way trading problems in [El-Yaniv et al. (2001) and Wang et al. (2016)]. The new approach
does not deal with the ratio directly and can have a similar level of complexity as the absolute regret analysis, as will be illustrated by applying it to a one-way trading problem.

Though there are many other robust criteria in the literature, none of them enjoys the combined advantages of tractability, smooth conservatism control, and the possibility of closed-form solutions. The Hurwicz criterion evaluates a solution by “a weighted sum of its worst and best possible outcomes.” Although the Hurwicz criterion often gives reasonable results, it could also lead to quite illogical answers, see Gaspars-Wieloch (2014) for a detailed analysis. Such inconsistency makes it difficult to manage and control the degree of conservatism reliably. Kalaı et al. (2012) suggest another criterion called the lexicographic robustness to reduce the impact of the primary role of the worst-case scenario to evaluate a solution, which faces a serious tractability challenge with a large uncertainty set. Roy (2010) proposes a criterion called the $bw$-robustness that requires an exact distribution to work. The $p$-robustness method by Snyder (2006) screens out by constraints those overly conservative solutions whose worst-case regret exceeds an upper limit, where either relative and absolute regret may be used. These constraints added for each scenario often makes the formulation intractable when they are numerous. It is also noted by Snyder (2006) that it can be very difficult to determine if a limit renders a problem feasible or not in some applications.

The contributions of this paper are as follows. I. The proposed ARM offers smooth control of conservatism for RO without tampering with the uncertainty set, which has great potential for practical applications. It can work with any specification of the uncertainty set and can be applied either independently or jointly with methods based on uncertainty sets, such as the uncertainty budget. II. Tractability results are obtained for ARM with commonly used types of linear RO problems. Such problems can be reformulated into traditional linear RO problems, for which tractable solution schemes and theoretical results have been actively developed. III. The properties of ARM are studied to facilitate its practical applications and the analysis of problems for closed-form solutions. Moreover, these theoretical properties lead to a new approach to competitive ratio analysis, which can help reduce the complexity of analysis for some problems. IV. The potential of ARM is demonstrated by applying it to the one-way trading problem. A closed-form solution is obtained, enabling the new approach to competitive ratio analysis to derive the ratio in a way much simpler than before. Numerical experiments are conducted to illustrate the continuous control of conservatism by ARM with some interesting insights.

The rest of this paper is arranged as follows. Section 2 provides formulations with the ARM criterion. The properties of ARM are studied in Section 3 to facilitate its applications, leading to a new approach to competitive ratio analysis. Then section 4 deals with the tractability of ARM, and in section 5 the well-studied robust one-way trading problem is employed to demonstrate the application of ARM. Finally, section 6 concludes with some future research suggestions.
2. Formulation. This section first introduces the ARM criterion with single-stage problems, then with multistage problems in which there is growing interest. In a single-stage problem, an action \( x \) is first taken from the set \( X \) of robustly feasible actions, then a scenario \( \zeta \in U \) is realized from the uncertainty set \( U \) of all scenarios considered, which may be a continuous or discrete set. The set \( X \) may also be continuous or discrete, for example, \( X = \{ x \in M : \forall \zeta \in U, g(x; \zeta) \leq 0 \} \), where \( g(x; \zeta) : X \times U \to \mathbb{R}^{n_\sigma} \), and \( M = \mathbb{R}^{n_r} \times \mathbb{Z}^{n_z} \) 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 \( \zeta \), and is given 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 observed \( \zeta \), indicating how potentially favorable a scenario is. Here the min and max operators are assumed well-defined, otherwise, they could be replaced by the inf and sup operators respectively.

The ARM criterion employs a continuous control parameter of conservatism \( \beta \in [0, \infty) \), and compares \( r(x, \zeta) \) with the \( \beta \)-adjusted benchmark \( \beta r^*(\zeta) \) to obtain an adjustable regret \( D(x; \zeta; \beta) = \beta r^*(\zeta) - r(x, \zeta) \). The worst-case regret \( \bar{D}(x; \beta) = \max_{\zeta \in U} D(x, \zeta; \beta) \) is minimized by the ARM criterion, which gives the ARM formulation:

\[
D(\beta) = \min_{x \in X} \bar{D}(x; \beta) = \min_{x \in X} \max_{\zeta \in U} \beta r^*(\zeta) - r(x, \zeta),
\]

\[
= \min_{x \in X} \max_{\zeta \in U} \beta \{ \max_{x' \in X} r(x', \zeta) \} - r(x, \zeta). \tag{1}
\]

The ARM criterion unifies a few well-known robust criteria into a continuum as \( \beta \) takes on different values. At \( \beta = 0 \), it degenerates into the maximin reward criterion. With \( \beta \) 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 \( \beta \) gets bigger, the ARM criterion gets more and more aggressive, according to the observations made by Lan et al. (2008) and Poursoltani and Delage (2021), which indicates that this \( \beta \) parameter can be continuously adjusted to moderate conservatism. Though the form in (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 purposely for moderating conservatism before.

It is helpful to intuitively explain how such moderation happens to capture the spirit. From the definition of \( \bar{D}(x; \beta) \) for a given \( x \), there is \( r(x, \zeta) \geq \beta r^*(\zeta) - D \) for any \( \zeta \in U \) if and only if \( D \geq \bar{D}(x; \beta) \). Taking \( r(x, \zeta) \) as a reward graph for \( x \) over the space \( U \), this inequality says that the reward graph is above the benchmark graph \( \beta r^*(\zeta) - D \). As \( D \) decreases and the benchmark graph rises, more and more reward graphs gets eliminated if a reward graph is too dissimilar to the benchmark in that it only has a low reward at a \( \zeta \) where the benchmark demands a high reward.
Thus when $D$ reaches $D(\beta)$, the final reward graph that remains above and gets selected is likely to be similar to the benchmark graph. As the $\beta$-adjusted benchmark graph becomes more aggressive for bigger $\beta$, demanding more for scenarios with higher potentials, the selected action is likely to be more aggressive as well. Hence, by increasing $\beta$, the ARM criterion may end up recommending more aggressive solutions. This intuitive explanation may lead to some theoretical developments, but the difficulty lies in appropriate definitions for the degree of conservatism and dissimilarity between reward graphs. More rigorous result on this for the one-way trading problem is given later in Corollary 3: the bigger the $\beta$, the more aggressive the optimal trading policy.

The single-stage formulation can be readily extended to multistage problems, where decisions are made sequentially as the uncertainty gradually reveals itself stage by stage. Let $t = 1, \ldots, T$ labels the stages, with a smaller $t$ for an earlier stage. The decision variable $x$ now consists of $T$ subvectors $(x_1, \ldots, x_T)$, with the stage decision $x_t$ for the decision in stage $t$. Likewise, a whole scenario $\zeta = (\zeta_1, \ldots, \zeta_T)$ now consists of stage scenarios $\zeta_t$ for each stage $t = 1, \ldots, T$.

To simplify discussions, it can be standardized without loss of generality to have action $x_t$ in each stage carried out before the realization of $\zeta_t$. For a stage scenario realized in the very beginning before any actions are taken, a dummy decision with only one choice of action (i.e. to participate in the decision process) can be prepended to the beginning to standardize the formulation.

Just as in multistage stochastic programming (MSP), it is implicitly assumed that the realization of scenarios is independent of decisions. The partial scenario $\zeta_{1:t} = (\zeta_1, \ldots, \zeta_{t-1})$ observed before stage $t$ defines all possible scenarios in the future, $\mathcal{U}(\zeta_{1:t}) = \{\zeta' : \zeta' \in \mathcal{U} : \zeta'_{1:t} = \zeta_{1:t}\}$. It is convenient to define the stage scenario set $\mathcal{U}_t(\zeta_{1:t}) = \{\zeta'_t : \zeta'_t \in \mathcal{U}(\zeta_{1:t})\}$.

Nonanticipativity in MSP requires that decisions are made without knowing the events not revealed yet, which is also honored here by having $x_t$ dependent only on what is known so far. Let $x_{1:t} = (x_1, \ldots, x_{t-1})$ 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 the current stage $t$ depends not only on $x_{1:t}$, but also on $\zeta_{1:t}$ through $\mathcal{U}(\zeta_{1:t})$. Therefore, let $X(h_t) = \{x' \in \mathcal{M} : x'_{1:t} = x_{1:t}, \forall \zeta' \in \mathcal{U}(\zeta_{1:t}), g(x'; \zeta') \leq 0\}$ be 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 stages or may be received at once in the end, so 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 $X(\zeta) = \{x | x_t \in X_t(x_{1:t}, \zeta_{1:t}), t = 1, \ldots, T\}$ is the set of all actions compatible with $\zeta$. By the end of the final stage, the complete history $h_{T+1} = (x, \zeta)$ is known, and the regret is readily computed as

$$D_T(h_{T+1}; \beta) = \beta r^*(\zeta) - r(x, \zeta). \quad (2)$$
The minimal worst-case regret in an earlier stage can be found by working backward from $D_T(h_{T+1}; \beta)$. Suppose $D_t(h_{t+1}; \beta)$ has already been worked out, the worst-case regret of taking action $x_t$ in the context of $h_t$ is

$$
\bar{D}_t(x_t, h_t; \beta) = \max_{\zeta \in D_t(h_t, 1)} D_t(h_{t+1}; \beta),
$$

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

$$
D_{t-1}(h_t; \beta) = \min_{x_t \in X_t(h_t)} \bar{D}_t(x_t, h_t; \beta),
$$

$$
= \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in D_t(h_t, 1)} D_t(h_{t+1}; \beta).
$$

The induction of $D_{t-1}(h_t; \beta)$ from $D_t(h_{t+1}; \beta)$ can be done recursively for $t = T, \cdots, 1$ backwards, working out a plain formulation that observes nonanticipativity. Note that when $t = 1$, there is no history in $h_1$, so let $D(\beta) = D_0(h_1; \beta)$ be the minimal worst-case regret for the whole problem.

An alternative formulation is based on policies, where a policy $\pi$ is a sequence of functions $\pi = \{\pi_t : \pi_t(h_t) \in X_t(h_t), t = 1, 2, \cdots, T\}$ so that a decision $x_t = \pi_t(h_t)$ is made for all possible history $h_t$. Clearly, nonanticipativity is automatically taken care of by this arrangement. 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 deterministic policies are focused on to simplify discussions. The regret for a policy $\pi$ in the end of the last period with a full history $h_{T+1} = (x, \zeta)$ is simply

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

Working backward from that, the worst-case regret for $t = T, \cdots, 1$ is found recursively as

$$
D^\pi_{t-1}(h_t; \beta) = \max_{\zeta_t \in U_t(\zeta_{1:t})} D^\pi_t(h_{t+1}; \beta),
$$

where $h^\pi_{t+1} = ((x_{1:t}, \pi_t(h_t)), (\zeta_{1:t}, \zeta_t))$ denote the history evolution under $\pi$. To compute the overall worst-case regret $D^\pi(\beta) = D^\pi_0(h_1; \beta)$ (since $h_1$ is empty), simply apply (6) recursively to have

$$
D^\pi(\beta) = \max_{\zeta_1 \in U_1(\zeta_{1,1})} D^\pi_1(h^\pi_2; \beta)
$$

$$
= \max_{\zeta_1 \in U_1(\zeta_{1,1})} \max_{\zeta_2 \in U_2(\zeta_{1,2})} D^\pi_2(h^\pi_3; \beta)
$$

$$
= \max_{\zeta_1 \in U_1(\zeta_{1,1})} \cdots \max_{\zeta_T \in U_T(\zeta_{1:T})} D^\pi_T(h^\pi_{T+1}; \beta)
$$

$$
= \max_{\zeta \in U} D^\pi_T(h^\pi_{T+1}; \beta).
$$

Let $\Pi$ be the set of all policies, and $r^*(\zeta) = r(\pi(\zeta), \zeta)$, where $x = \pi(\zeta)$ is the whole sequence of stage decisions for scenario $\zeta$ by policy $\pi$. The policy-based formulation is given by

$$
\min_{\pi \in \Pi} D^\pi(\beta) = \min_{\pi \in \Pi} \max_{\zeta \in U} \beta r^*(\zeta) - r^*(\zeta).
$$
3. Properties. The properties of ARM are studied in this section to facilitate its application and the development of a new approach to competitive ratio analysis. The correspondence 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

\[
D_{t-1}(h_t; \beta) = D^*_{t-1}(h_t; \beta), \text{ for } t = 1, \cdots, T + 1,
\]

with an optimal policy \( \pi^* \) constructed by

\[
\pi^*_t(h_t) = \arg\min_{x_t \in X_t(h_t)} \max_{\zeta_t \in U_t(h_t)} D_t(h_{t+1}; \beta), \text{ for } t = 1, \cdots, T,
\]

where the \( \arg\min \) 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

\[
D_{t-1}(h_t; \beta) = \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in U_t(h_t)} D_t(h_{t+1}; \beta)
= \max_{\zeta_t \in U_t(h_t)} D^*_{t}(h^*_{t+1}; \beta)
= D^*_{t-1}(h_t; \beta),
\]

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

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

\[
D_{t-1}(h_t; \beta) \leq D^*_{t-1}(h_t; \beta), \text{ for } t = 1, 2, \cdots, T + 1
\]

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 \( t + 1 \): \( D_t(h_{t+1}; \beta) \leq D^*_{t}(h_{t+1}; \beta) \), then show (11) also holds for \( t \). Recall (4) and replace \( D_t(h_{t+1}; \beta) \) with \( D^*_{t}(h_{t+1}; \beta) \) to have

\[
D_{t-1}(h_t; \beta) \leq \min_{x_t \in X_t(h_t)} \max_{\zeta_t \in U_t(h_t)} D^*_{t}(h^*_{t+1}; \beta)
\leq \max_{\zeta_t \in U_t(h_t)} D^*_{t}(h^*_{t+1}; \beta)
= D^*_{t-1}(h_t; \beta),
\]

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 \( \pi^* \) 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 \( D^*_{0}(\beta) = D(\beta) \) by (9) with \( t = 1 \). Besides this direct proof, an alternative proof of...
Theorem \ref{thm:interchangeability} can be made by recursively applying the interchangeability principle of \cite{Shapiro2017}, 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 \cite{Shapiro2017} 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).

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

\textbf{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 $\beta$ according to \cite{Shapiro2017}, which completes the initial step. For the induction step, show that if $D_t(h_{t+1}; \beta)$ is continuous in $\beta$, then so is $D_{t-1}(h_t; \beta)$. It is clear that $D_t(h_t; \beta)$ is continuous in $\beta$ as it is a point-wise max of continuous functions by \cite{Shapiro2017}. Likewise, $D_{t-1}(h_t; \beta)$ is also continuous with regard to $\beta$ by \cite{Shapiro2017}, which completes the proof. \hfill \blacksquare

\textbf{Theorem 2} For $0 \leq \beta_1 < \beta_2$, let $\pi_1^*, i \in \{1,2\}$ be an optimal policy when $\beta = \beta_i$, and $\zeta_{ij} = \arg \max_{\zeta \in U} \beta_i r^\pi(\zeta) - r_1^\pi(\zeta), i,j \in \{1,2\}$, then there is

$$r^\pi(\zeta_{21}) \geq \frac{D(\beta_2) - D(\beta_1)}{\beta_2 - \beta_1} \geq r^\pi(\zeta_{12}). \quad (12)$$

\textbf{Proof:} By the definition of $\pi_2^*$ and $\zeta_{12}^*$, as well as Theorem \ref{thm:interchangeability}, there is

$$D(\beta_1) = \min_{\pi \in \Pi} \max_{\zeta \in U} \beta_1 r^\pi(\zeta) - r^\pi(\zeta) \leq \max_{\zeta \in U} \beta_1 r^\pi(\zeta) - r_2^\pi(\zeta) = \beta_1 r^\pi(\zeta_{12}^*) - r_2^\pi(\zeta_{12}^*).$$

And there is $D(\beta_2) = \max_{\zeta \in U} \beta_2 r^\pi(\zeta) - r_2^\pi(\zeta) \geq \beta_2 r^\pi(\zeta_{12}^*) - r_2^\pi(\zeta_{12}^*)$. Therefore $D(\beta_2) - D(\beta_1) \geq (\beta_2 - \beta_1) r^\pi(\zeta_{12}^*)$. Similarly,

$$D(\beta_2) = \min_{\pi \in \Pi} \max_{\zeta \in U} \beta_2 r^\pi(\zeta) - r^\pi(\zeta) \leq \max_{\zeta \in U} \beta_2 r^\pi(\zeta) - r_1^\pi(\zeta) = \beta_2 r^\pi(\zeta_{21}^*) - r_1^\pi(\zeta_{21}^*).$$

And there is $D(\beta_1) = \max_{\zeta \in U} \beta_1 r^\pi(\zeta) - r_1^\pi(\zeta) \geq \beta_1 r^\pi(\zeta_{21}^*) - r_1^\pi(\zeta_{21}^*)$. Thus $D(\beta_2) - D(\beta_1) \leq (\beta_2 - \beta_1) r^\pi(\zeta_{21}^*)$. Therefore, \cite{Shapiro2017} follows immediately. \hfill \blacksquare

The continuity of Lemma \ref{thm:interchangeability} is a basic property useful for other analytical results. Note that when $\forall \zeta, r^\pi(\zeta) > 0$, Theorem \ref{thm:interchangeability} ensures the monotonicity of $D(\beta)$. 
3.1. Convexity. The convexity of $D(\beta)$ is studied next. Note that $\beta r^*(\zeta) - r^*(\zeta)$ is linear in $\beta$, thus the function

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

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

**Lemma 2** A continuous function $f(y)$ with 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 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 $0 \leq k_1 < \lambda < k_2 \leq 1$, $g(k_1) = g(k_2) = 0$ 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 $f(y_1') = k_i f(y_1) + (1 - k_i) f(y_2), i = 1, 2$ 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 $\pi$ dominates another policy $\pi'$ (denoted as $\pi \succeq \pi'$) if for all $\zeta \in \mathcal{U}$ there is $r^*(\zeta) \geq r^*(\zeta)$. Similarly, a scenario $\zeta$ dominates another scenario $\zeta'$ (denoted as $\zeta \succeq \zeta'$) if for all $\pi \in \Pi$ there is $r^*(\zeta) \leq r^*(\zeta')$. According to Lan et al. (2008), dominated policies and scenarios can be eliminated in a similar way as iterated elimination of dominated strategies in game theory.

**Definition 1 (Reward Convexity (RC))** The set $\Pi$ has the property of RC if there is

$$\forall \pi_1, \pi_2 \in \Pi \exists \pi \in \Pi \exists \lambda \in (0, 1) \, \forall \zeta \in \mathcal{U} \, r^*(\zeta) = \lambda r^{*1}(\zeta) + (1 - \lambda)r^{*2}(\zeta). \quad (14)$$

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

**Definition 2 (Dominance Convexity (DC))** The set $\Pi$ has the DC property if there is

$$\forall \pi_1, \pi_2 \in \Pi \exists \pi \in \Pi \exists \lambda \in (0, 1) \, \forall \zeta \in \mathcal{U} \, r^*(\zeta) \geq \lambda r^{*1}(\zeta) + (1 - \lambda)r^{*2}(\zeta). \quad (15)$$
Clearly, if $\Pi$ 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 remain 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 $\Pi$, so that any $\pi \in \Pi$ is dominated by a $\hat{\pi} \in \hat{\Pi}$. Let $\Pi^\star(\beta)$ be the set of all optimal policies for a given $\beta$.

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

**Proof:** (i.a) from $\Pi$ to $\Pi^\star(\beta)$. 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 $\hat{\pi} \succeq \pi$, 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 $\hat{\pi}_1 \succeq \pi_1, \hat{\pi}_2 \succeq \pi_2$. There exists $\hat{\pi} \in \hat{\Pi}$ and $\lambda \in (0, 1)$ such that $r^{\hat{\pi}}(\zeta) \geq \lambda r^{\hat{\pi}_1}(\zeta) + (1 - \lambda)r^{\hat{\pi}_2}(\zeta)$. As $\hat{\pi} \in \Pi$, hence $\Pi$ also has the DC property.

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

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

Now show that $\pi^\prime \in \Pi^\star(\beta)$ as follows:

$$
D(\beta) = \min_{\pi \in \Pi} \max_{\zeta \in \mathcal{U}} \beta r^\pi(\zeta) - r^\pi(\zeta) \\
\leq \max_{\zeta \in \mathcal{U}} \beta r^\pi(\zeta) - r^{\pi^\prime}(\zeta) = D^{\pi^\prime}(\beta) \\
\leq \max_{\zeta \in \mathcal{U}} \beta r^\pi(\zeta) - \left( \lambda r^{\pi_1^\star}(\zeta) + (1 - \lambda)r^{\pi_2^\star}(\zeta) \right) \\
\leq \lambda \left( \max_{\zeta \in \mathcal{U}} \beta r^\pi(\zeta) - r^{\pi_1^\star}(\zeta) \right) + \\
(1 - \lambda) \left( \max_{\zeta \in \mathcal{U}} \beta r^\pi(\zeta) - r^{\pi_2^\star}(\zeta) \right) \\
= \lambda D(\beta) + (1 - \lambda)D(\beta) = D(\beta).
$$

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

Theorem 3 is useful to prove DC property for $\Pi$ by simply focusing on the non-dominated subset $\hat{\Pi}$. Meanwhile, the DC property of $\Pi^\star(\beta)$ may help select a more preferable policy when there are many optimal policies.
Theorem 4 If $\Pi$ has the DC property, then $D(\beta)$ is convex in $\beta$.

Proof: Let $\pi_1^*$ be an optimal policy for $\beta_i, i = 1, 2$. By (15) there exists $\pi' \in \Pi$ such that $\exists \lambda \in (0, 1) \forall \zeta \in U \ r^\pi_1(\zeta) \geq \lambda r^\pi_1(\zeta) + (1 - \lambda) r^\pi_2(\zeta)$. Let $\beta = \lambda \beta_1 + (1 - \lambda) \beta_2$ and proceed as follows:

$$D(\beta) = \min_{\pi \in \Pi} \max_{\zeta \in U} \beta_1 r^\pi_1(\zeta) - r^\pi_1(\zeta)$$

$$\leq \max_{\zeta \in U} \beta_1 r^\pi_1(\zeta) - r^\pi_1'(\zeta)$$

$$\leq \max_{\zeta \in U} \beta_1 r^\pi_1(\zeta) - \left( \lambda r^\pi_1(\zeta) + (1 - \lambda) r^\pi_2(\zeta) \right)$$

$$\leq \lambda \left( \max_{\zeta \in U} \beta_1 r^\pi_1(\zeta) - r^\pi_1(\zeta) \right) +$$

$$(1 - \lambda) \left( \max_{\zeta \in U} \beta_2 r^\pi_1(\zeta) - r^\pi_2(\zeta) \right)$$

$$= \lambda D(\beta_1) + (1 - \lambda) D(\beta_2).$$

Therefore $D(\beta)$ is convex in $\beta$ by Lemma 2.

The DC property is not hard to find, such as in most linear robust optimization problems, hence Theorem 4 readily applies. As the benchmark is linearly scaled up by $\beta$, the convexity of $D(\beta)$ indicates that as $\beta$ increases, the corresponding optimal solution is less and less speedy to catch up with the scaling up of the benchmark. Meanwhile, the convexity may help design faster algorithms than binary searches in Averbakh (2005) that numerically compute competitive ratios.

3.2. Competitive Ratio. In reward maximization applications, the competitive ratio can be defined as

$$\max_{\pi \in \Pi} \min_{\zeta \in U} \frac{r_\pi(\zeta)}{r_\pi^*(\zeta)},$$

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

Lemma 3 Assume $r^*(\zeta) > 0$ for all $\zeta \in U$, the $\beta_0$ 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 $\pi^*$ is an optimal solution to (16), and vice versa. By Theorem 1 and 3 there is

$$\left\{ \begin{array}{l}
0 = \min_{\pi \in \Pi} \max_{\zeta \in U} \beta_0 r^\pi_1(\zeta) - r^\pi_1(\zeta) \\
\pi^* \in \arg\min_{\pi \in \Pi} \max_{\zeta \in U} \beta^* r^\pi_1(\zeta) - r^\pi_1(\zeta)
\end{array} \right\}$$

$$\Leftrightarrow \left\{ \begin{array}{l}
0 = \max_{\zeta \in U} \beta_0 r^\pi_1(\zeta) - r^\pi_1(\zeta) \\
\forall \pi \in \Pi \ 0 \leq \max_{\zeta \in U} \beta_0 r^\pi_1(\zeta) - r^\pi_1(\zeta)
\end{array} \right\}$$
\[
\begin{aligned}
&\exists \zeta \in \mathcal{U} 0 = \beta_0 r^*(\zeta) - r^\pi(\zeta) \\
&\forall \zeta \in \mathcal{U} 0 \geq \beta_0 r^*(\zeta) - r^\pi(\zeta) \\
&\forall \pi \in \Pi \exists \zeta \in \mathcal{U} 0 \leq \beta_0 r^*(\zeta) - r^\pi(\zeta) \\
&\exists \zeta \in \mathcal{U} : \beta_0 = r^\pi(\zeta)/r^*(\zeta) \\
&\forall \zeta \in \mathcal{U} : \beta_0 \leq r^\pi(\zeta)/r^*(\zeta) \\
&\forall \pi \in \Pi \exists \zeta \in \mathcal{U} \beta_0 \geq r^\pi(\zeta)/r^*(\zeta)
\end{aligned}
\]

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

Based on the result of Lemma 3, the next lemma gives the condition for the existence of a unique competitive ratio.

**Lemma 4** 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:

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

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

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

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 \(\beta\). 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 1 and Theorem 2.

Variants of Lemma 3 and 4 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):

\[
\max_{x \in X, y(\cdot) \in \mathcal{U}} \inf_{\zeta \in \mathcal{U}} (C \zeta + c)^T x + d^T y(\zeta) + f^T \zeta,
\]

\[
s.t. \ Ax + By(\zeta) \leq \Psi \zeta + \psi, \forall \zeta \in \mathcal{U},
\]

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 \(\zeta\) has been revealed.
The constants are \( \Psi \in \mathbb{R}^{m \times n \times n} \), \( C \in \mathbb{R}^{nx \times n} \), \( c \in \mathbb{R}^{nx} \), \( d \in \mathbb{R}^{nx} \), \( f \in \mathbb{R}^{n \times n \times n} \), \( A \in \mathbb{R}^{m \times nx} \) and \( B \in \mathbb{R}^{m \times ny} \). Both \( \mathcal{X} \) and \( \mathcal{U} \) are nonempty polyhedra: \( \mathcal{X} := \{ x \in \mathbb{R}^{nx} | Wx \leq v \} \) with \( W \in \mathbb{R}^{r \times nx} \) and \( v \in \mathbb{R}^{r} \), and \( \mathcal{U} := \{ \zeta \in \mathbb{R}^{nz} | P\zeta \leq q \} \) 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

\[
h(x, \zeta) := \sup_y (C\zeta + c)^T x + d^T y + f^T \zeta
\]

\[
s.t.: \quad Ax + By \leq \Psi(x)\zeta + \psi.
\]

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

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

\[
\min_{x \in \mathcal{X}} \sup_{x' \in \mathcal{X}} \beta h(x', \zeta) - h(x, \zeta),
\]

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 $(x, \zeta, y)$ 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,

$$\forall x \in \mathcal{X}, \forall \zeta \in \mathcal{U}, \exists y \in \mathbb{R}^n, Ax + By \leq \Psi(x)\zeta + \psi.$$ (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 $\bar{\zeta}: \mathcal{X} \rightarrow \mathcal{U}$ 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:

$$\forall x \in \mathcal{X}, \inf_{\zeta \in \mathcal{U}} h(x, \zeta) > -\infty.$$ (22)

Assumption 5 (Finite Best-Case) There is an upper bound on the best-case profit achievable:

$$\sup_{x \in \mathcal{X}, \zeta \in \mathcal{U}} h(x, \zeta) < \infty.$$ (22)

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,

$$h(x, \zeta) := \sup_y c^T x + d^T y$$

subject to:

$$Ax + By \leq \Psi(x)\zeta + \psi.$$ (23)

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

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

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

where $\zeta' = [\zeta^T \ x'^T \ y'^T]^T$ is called a lifting of $\zeta$ 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 5, the LROARM problem with right-hand side uncertainty is equivalent to the following TSLRO/FR problem:

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

subject to:

$$Ax + By(\zeta') \leq \Psi' \zeta' + \psi, \forall \zeta' \in \mathcal{U}'$$.
where $\zeta' = [c^Ta' + T^Ty]' \in \mathbb{R}^{n_x + n_y + n_y}$, $y: \mathbb{R}^{n_x + n_y} \rightarrow \mathbb{R}^{n_y}$, $f' := [0^T - c^T - d^T]'$, $\Psi' := [\Psi 0 0]$, and the lifted uncertainty set $U'$ can be defined by

$$U' := \{ \zeta' \in \mathbb{R}^{n_x + n_y + n_y} | P' \zeta' \leq q' \}, \quad \text{with}$$

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

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 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:

$$\text{LROARM} \equiv \min_{x \in \mathcal{X}} \sup_{\zeta \in U} \left\{ \sup_{x' \in \mathcal{X}, y' \in \mathcal{Y}(x', \zeta)} \beta(c^T x' + d^T y') - \sup_{y \in \mathcal{Y}(x, \zeta)} c^T x + d^T y \right\},$$

$$\equiv \min_{x \in \mathcal{X}, \zeta \in U} \inf_{x' \in \mathcal{X}, y' \in \mathcal{Y}(x', \zeta)} \sup_{y \in \mathcal{Y}(x, \zeta)} \beta(c^T x' + d^T y') - (c^T x + d^T y),$$

$$\equiv -\max_{x \in \mathcal{X}, \zeta \in U} \inf_{x' \in \mathcal{X}, y' \in \mathcal{Y}(x', \zeta)} \sup_{y \in \mathcal{Y}(x, \zeta)} c^T x + d^T y \beta(c^T x' + d^T y'),$$

$$\equiv -\max_{x \in \mathcal{X}, y' \in \mathcal{Y}(x', \zeta')} \inf_{x' \in \mathcal{X}, \zeta' \in \mathcal{U}'} c^T x + d^T y(\zeta') + f'^T \zeta',$$

s.t. $Ax + By(\zeta') \leq \Psi(\zeta' + \psi), \forall \zeta' \in U'$,

where $Y(x, \zeta) := \{ y \in \mathbb{R}^{n_y} | Ax + B y \leq \Psi \zeta + \psi \}$, 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 $\bar{x} \in \mathcal{X}, \bar{\zeta} \in U$, and bary $Y(\bar{x}, \bar{\zeta})$. It is clear that $\zeta' := [c^T \bar{x}^T \bar{y}^T]' \in U'$ so that the triplet $\bar{x}, \bar{\zeta}, \bar{y}$ satisfies Assumption 1 for the reformulation. Therefore, Assumption 1 carries through naturally. Second, since the feasible set for the recourse problem is the same in LROARM and its TSLRO/FR reformulation, Assumption 2 carries also carries through unscathed. Third, one can show that Assumption 3 also carries through if Assumption 2 holds. Simply let $\bar{\zeta}: \mathcal{X} \rightarrow U$ map an $x$ to a $\zeta$ that must exist according to Assumption 2 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 $\zeta'(x) := [c^T(x) x'^T \bar{y}' \bar{y}]'$ provides that special $\zeta'$ 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 2 and 3 satisfied by the LROARM problem, let the recourse problem that appears in the TSLRO/FR reformulation as $h'(x, \zeta')$,

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

4.2. Objective Uncertainty

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

\[
h(x, \zeta) := \sup_y c^T x + d^T (\zeta)y
\]

\[\text{s.t. : } Ax + By \leq \psi.
\]

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

\[
h(x, \zeta) := \sup_y x, y \in X \in Y c^T (\zeta)y + d^T (\zeta)y
\]

\[\text{s.t. : } y = x
\]

\[Ax + By \leq \psi.
\]

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

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

\[\text{s.t. } A'x + B'y'(\zeta') \leq \Psi' \zeta' + \psi',
\]

where \( y' : \mathbb{R}^{n+m} \rightarrow \mathbb{R}^{m+r} \), whereas \( \mathcal{U}' \) is the new uncertainty set defined as:

\[
\mathcal{U}' := \{ \zeta' \in \mathbb{R}^{n+m} | P' \zeta' \leq q' \}, \text{ with}
\]

\[
P' = \begin{bmatrix}
P & 0 \\
-D & B^T \\
D & -B^T
\end{bmatrix}
\]

and \( q' = \begin{bmatrix}
q \\
d
-d
\end{bmatrix} \).

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:

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

\[
\text{s.t. } Ax' + By' \leq \psi,
\]

\[
Wx' \leq v,
\]

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

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

\[
\text{s.t. } A^T \lambda + W^T \gamma = c,
\]

\[
B^T \lambda = d(\zeta),
\]

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:

\[
\min \sup_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}} \beta h(x', \zeta) - h(x, \zeta),
\]

\[
\equiv \min \sup_{x \in \mathcal{X}} \left\{ \sup_{\zeta \in \mathcal{U}} \beta h(x', \zeta) - \inf_{\rho \in \Phi(\zeta)} c^T x + (\psi - Ax)^T \rho \right\},
\]

\[
\equiv \min \sup_{x \in \mathcal{X}} \sup_{\zeta \in \mathcal{U}, \rho \in \Phi(\zeta)} \beta h(x', \zeta) - (c^T x + (\psi - Ax)^T \rho),
\]

\[
\equiv \min \inf_{x \in \mathcal{X}, \zeta \in \mathcal{U}, \rho \in \Phi(\zeta), (\lambda, \gamma) \in \Lambda(\zeta)} \beta (\psi^T \lambda + v^T \gamma) - (c^T x + (\psi - Ax)^T \rho),
\]

\[
\equiv -\max \inf_{x \in \mathcal{X}, \zeta \in \mathcal{U}, \rho \in \Phi(\zeta), (\lambda, \gamma) \in \Lambda(\zeta)} c^T x + (\psi - Ax)^T \rho - \beta (\psi^T \lambda + v^T \gamma),
\]

where \(\Lambda(\zeta) := \{(\lambda, \gamma) \in \mathbb{R}^m \times \mathbb{R}^r| (32), (53), \lambda, \gamma \geq 0\}\) and \(\Phi(\zeta) := \{\rho \in \mathbb{R}^m| B^T \rho = d(\zeta), \rho \geq 0\}\). By having \(\zeta' := [\xi^T \rho^T]^T\) 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 \(h(x, \tilde{\zeta}(x)) < \infty\). This implies by LP duality that there must be a feasible \(\tilde{\rho} \in \Phi(\tilde{\zeta}(x))\). Moreover, Assumption 5 implies that \(\sup_{x \in \mathcal{X}} h(x', \tilde{\zeta}(x)) < \infty\) hence once again LP duality ensures that there exists a pair \((\tilde{\lambda}, \tilde{\gamma}) \in \Lambda(\tilde{\zeta}(x))\). The TSLRO/FR reformulation therefore satisfies Assumption 1 using \((x, \tilde{\zeta}(x), \tilde{\rho}, \tilde{\lambda}, \tilde{\gamma})\). Next, the fact that the TSLRO/FR reformulation satisfies Assumption 2 follows similarly from imposing Assumption 3 on LROARM because the existence of a pair \((\tilde{\lambda}, \tilde{\gamma}) \in \Lambda(\zeta)\) 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,

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

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. There has been considerable work on the one-way trading problem, using tools from stochastic optimization, dynamic programming and reinforcement learning. The first work to study it in a RO formulation with the relative regret or competitive ratio criterion is El-Yaniv et al. (2001), which has received significant attention in the literature. A follow-up study by Wang et al. (2016) employs the minimax regret criterion and obtains closed-form solutions. The purpose to study it with the ARM criterion is three-fold. First, to show through backward induction that ARM formulations can have closed-form solutions, while recovering the main results of both El-Yaniv et al. (2001) and Wang et al. (2016) by setting $\beta$ to a proper value. Second, to apply the newly proposed approach to competitive ratio analysis, without depending on acute intuition and special insights as in El-Yaniv et al. (2001). Note that the difficulty of the analysis is comparable to that in Wang et al. (2016), and much easier than that in El-Yaniv et al. (2001). Third, to prove theoretically that the policy indeed becomes more aggressive as $\beta$ increases, and to show how $\beta$ affects the performance via numerical simulations. The one-way trading problem serves to demonstrate the potential of ARM.
5.1. Problem Formulation. Consider the one-way trading problem to sell a total 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 \(t = 1, \ldots, T\). The trader is a price-taker and must decide on the amount \(x_t\) to sell at the current price \(p_t\) in each period without knowing the future prices. The goal is to maximize the total sales revenue in the end.

It is helpful to adopt the notations in section 2. A scenario \(\zeta\) here simply corresponds to the prices \(p = (p_1, \ldots, p_T)\) revealed over time, with \(\zeta_t = p_t\). As the prices are independent of each other, there is \(\mathcal{U}(p_{1:T}) = [m, M]^{T}\), and \(\mathcal{U} = [m, M]^T\). Without loss of generality, the total amount of goods is one unit, and the action is \(x = (x_1, \ldots, x_T)\) 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 \(h_t = (x_{1:t}, p_{1:t})\) and \(q_t = 1 - \sum_{s=1}^{t-1} x_s\) is the remaining amount to sell given \(h_t\), but in the last period there is \(X_T(h_T) = [q_T, q_T]\) as nothing should remain. The reward is accumulated over time, so let \(r_t = \sum_{s=1}^{t-1} p_s x_s\) be the rewards accumulated for \(h_t\), the reward in the end is \(r(x, p) = r_{T+1}\). Let \(\hat{p}_t = \max\{p_s : s = 1, \ldots, t-1\}\) denote the highest price seen in \(h_t\), then \(r^*(p) = \max\{r(x, p) : \sum_{t=1}^{T} x_t = 1\} = \hat{p}_{T+1}\). At the end of the last stage \((2)\) becomes

\[
D_T(h_{T+1}; \beta) = \beta \hat{p}_{T+1} - r_{T+1}.
\]

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 \((2)\):

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

but the difference is 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 \([m, M]\),

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

where \(y^{+j} = \max(0, y)^j\) denote the positive part of \(y\) raised to the \(j^{th}\) power. Let \(P_j^{-1}(y) = q\) be the inverse of \(y = P_j(q)\) for \(q \in [0, \beta j]\).

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

\[
= \max(\beta \hat{p}_T, \beta M - (M - m) q_T) - R_T
\]

\[
= \beta \max(\hat{p}_T, P_1(q_T)) - R_T,
\]
where $R_t = r_t + mq_t$ for $t = 1, \ldots, T$ is the lower bound on $r_{T+1}$ given $h_t$. Note that the special but trivial case of $\beta = 0$ is not included here. Continue on with (36) for $t = T - 1, \ldots, 1$, 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 $t = 1, 2, \ldots, T$ is

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

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

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

**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

$$\bar{D}_t(h_t, p_t; \beta) = \min_{x_t \in X_t(h_t)} D_t(h_{t+1}; \beta)$$

$$= \min_{q_{t+1} \in [0, q_t]} \beta \max(\hat{p}_{t+1}, P_n(q_{t+1})) - R_{t+1}, \tag{39}$$

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 $\bar{p}_i = \max(\hat{p}_i, P_n(q_i)) \in [m, M]$, and from (36) there is
\[
D_{t-1}(h_t; \beta) = \max_{p_t \in [m, M]} D_t(h_t, p_t; \beta) = \max \left( \max_{p_t \in [\bar{p}_i, M]} \frac{D_t(h_t, p_t; \beta)}{D_t(h_t, p_t; \beta)} \right)
\]  
(43)

For the branch with $p_t \in [m, \bar{p}_i]$ in (33), consider two cases: (i) $\bar{p}_i = \hat{p}_i \geq P_n(q_t)$ and (ii) $\bar{p}_i = P_n(q_t) > \hat{p}_i$. In case (i) there is $\pi_{t+1} = \max(\hat{p}_i, p_t) = \hat{p}_i \geq P_n(q_t)$, therefore $P_n(\hat{p}_{t+1}) \leq q_t$ and (38) simplifies to $q^*_{t+1} = P_n^{-}(\hat{p}_{t+1})$, thus $P_n(q^*_{t+1}) = \hat{p}_{t+1} = \bar{p}_i$. In case (ii) there is $\pi_{t+1} \leq P_n(q_t)$, therefore $P_n^{-}(\bar{p}_{t+1}) \leq q_t$ and (38) simplifies to $q^*_{t+1} = q_t$, thus $P_n(q^*_{t+1}) = P_n(q_t) = \bar{p}_i$. As in both cases there is $P_n(q^*_{t+1}) = \bar{p}_i$, from (12) there is $D_t(h_t, p_t; \beta) = \beta \hat{p}_t - (r_{t+1} + mq^*_{t+1}) = \beta \hat{p}_t - r_t - p_t x_t - mq^*_{t+1}$, 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
\[
\max_{p_t \in [m, \bar{p}_i]} \tilde{D}_t(h_t, p_t; \beta) = \beta \hat{p}_t - r_t - mq_t = \beta \bar{p}_t - R_t.
\]

For the branch in (43) with $p_t \in [\bar{p}_i, M]$, as $p_t \geq \hat{p}_i \geq \bar{p}_i$, there is $\pi_{t+1} = p_t \geq \hat{p}_i \geq P_n(q_t)$, thus $P_n^{-}(\bar{p}_{t+1}) \leq q_t$ and (38) simplifies to $q^*_{t+1} = P_n^{-}(\hat{p}_{t+1})$. Therefore $P_n(q^*_{t+1}) = \bar{p}_{t+1} = p_t$, and (42) simplifies to $D_t(h_t, p_t; \beta) = \beta p_t - r_t - p_t x_t^* - mq^*_{t+1} = \beta p_t - r_t - p_t(q_t - q^*_{t+1}) - mq^*_{t+1} = (\beta - q_t + q^*_{t+1}) p_t - mq^*_{t+1} - r_t = (\beta - q_t + q^*_{t+1}) P_n(q^*_{t+1}) - mq^*_{t+1} - r_t$. Now consider this function
\[
d(z) = (\beta - q_t + z) P_n(z) - m z - r_t, z \in [0, 1].
\]
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 $z^* = q_t/(n+1)$ 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 $\tilde{D}_t(h_t, p_t; \beta) = d(P_n^{-}(p_t))$ for $p_t \in [\bar{p}_i, M]$, consider two cases. Case (i) $P_n(z^*) \geq \bar{p}_i$. As $P_n^{-}(M) = 0 \leq z^* \leq P_n^{-}(\bar{p}_i)$, there is $\max_{p_t \in [\bar{p}_i, M]} \tilde{D}_t(h_t, p_t; \beta) = d(z^*)$. Thus according to (43) there is
\[
D_{t-1}(h_t; \beta) = \max(\beta \bar{p}_t - R_t, d(z^*)).
\]

Case (ii) $P_n(z^*) < \bar{p}_i$. As $q_t \geq z^* \geq P_n^{-}(\bar{p}_i)$, there is
\[
\max_{p_t \in [\bar{p}_i, M]} \tilde{D}_t(h_t, p_t; \beta) = \max_{p_t \in [\bar{p}_i, M]} d(P_n^{-}(p_t)) = \max_{z \in [0, P_n^{-}(\bar{p}_i)]} 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

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

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

\textbf{Corollary 1} The minimal worst-case regret \( D(\beta) \) for the one-way trading problem is a convex function of \( \beta \):

\[
  D(\beta) = \beta(M - m) \left( 1 - \frac{1}{\beta T} \right)^T - (1 - \beta)m,
\]

\textit{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 of analysis 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). Such simplification is impossible with the analysis and results in Wang et al. (2016).

\textbf{Corollary 2} The competitive ratio defined in (16) for the one-way trading problem is the unique root \( \beta_0 \) of \( D(\beta) \) as defined in (44).

\textit{Proof:} As \( r^*(\zeta) \geq m > 0 \), it follows from Lemma 3 and Lemma 4.

This is in perfect agreement with El-Yaniv et al. (2001), except that they define competitive ratio as the inverse of \( \beta_0 \). 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 $\beta$ 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 

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

increases in $\beta$, therefore $q^*_{t+1}$ increases as $\beta$ increases. 

This corollary shows that the optimal policy gets more and more aggressive as $\beta$ increases, which illustrates the continuous moderation of conservatism by $\beta$ in this problem.

5.3. Numerical Study. The effect of $\beta$ on the one-way trading policy is further demonstrated by numerical simulations next. The prices are I.I.D. with a uniform distribution on $[1, 2]$ for all the $T = 5$ periods, but in the ARM formulation only the price bounds $m = 1, M = 2$ are given. For each of the 300 evenly spaced values of $\beta \in (0, 3]$, the policy from (38) is executed on the same sequence of randomly generated prices and the overall revenue is accrued over all five periods. Repeat this process $N = 10,000$ times and the average and standard deviation of the revenue for each $\beta$ 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 as $\beta$ increases. The first phase ($\beta < 0.8$) witnesses increases in the average overall revenue with decreases in the standard deviation as $\beta$ increases, indicating that an overly conservative policy can not only hurt the revenue, but it may not really reduce the overall risk as it is supposed to. An explanation is that at $\beta = 0$ the policy trades everything in the first period at any price, then as $\beta$ increases the policy trades less at lower prices and reserves more for an opportunity to sell at higher prices in the future, which not only increases expected overall 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 risk. The third phase ($\beta > 1.5$) experiences decreasing revenue while the risk increases. This is probably 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. If there are more periods to sell, it pays to be more aggressive, and one would expect that the transition to the next phase will come later on $\beta$, which is indeed the case as in Fig. 2 with $T = 15$.

These numerical experiments provide some interesting insights: 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 can trade-off and find a sweet spot in between, with the optimal $\beta$ depending on the context. These insights can be gleaned from the following observations. Firstly, an overly conservative robust policy such as obtained by the maximin reward criterion (i.e. $\beta = 0$) can suffer both lower performance and higher overall risk. 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 have the historic prices fixed and only the future prices randomized. It is helpful to point out that more conservative policies with smaller $\beta$ values indeed have lower conditional risks. Secondly, an overly aggressive robust policy for a big $\beta$ may also hurt the performance as it boldly exposes to future price risks by reserving too much, ending up selling a lot in the last period at any price. Last but not least, there is a trade-off between the two extremes by adjusting the $\beta$ value of the ARM criterion. As illustrated by the two settings with $T = 5$ and $T = 15$ respectively, the fine-tuned $\beta$ value depends on the problem context, which may well go beyond $\beta = 1$ for the minimax absolute regret criterion.

6. Conclusion. This paper proposes the ARM criterion with a control parameter ($\beta$) for continuous control of conservatism. By minimizing the worst-case adjustable regret, the ARM criterion chooses a solution that is likely to match the benchmark adjustable by the control parameter, which enables the moderation of conservatism of the recommended solution. Various theoretical properties of ARM are studied, such as continuity, monotonicity, and convexity, which may facilitate the analysis of problems to find closed-form solutions or design better numerical algorithms to calculate competitive ratios. These theoretical investigations also lead to a new approach for competitive ratio analysis, which turns out to be much simpler than the traditional approach in the analysis of the one-way trading problem.

The tractability of ARM is studied for two-stage linear problems. 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, which produces a closed-form solution and recovers the results for both the competitive ratio and the minimax regret analysis. Analysis of the closed-form solution reveals that the optimal policy indeed gets more and more aggressive at the control parameter increases. A numerical study is carried out to demonstrate the effects of smooth control of conservatism in the one-way trading problem. Some insights are gleaned from the numerical study: the extremes of being totally conservative or aggressive may suffer from both lower average reward and higher overall risk at the same time, the proper $\beta$ value is somewhere in between, which depends on the problem context.

The investigations of the ARM criterion carried out in this paper can hopefully serve as a starting point for future research. A rigorous analysis on how $\beta$ moderates conservatism is certainly of great theoretical interest and challenging at the same time. Conceivably, how to choose an appropriate $\beta$ value in real applications depends on the application context, and it is another worthy topic of interest for future study.
References

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

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

Averbakh I, 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, Nemirovski A (2009) Robust optimization (Princeton university press).

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

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

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

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

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

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

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

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

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

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

Delage E, Iancu DA (2015) Robust multistage decision making. The operations research revolution, 20–46 (INFORMS).

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

El-Yaniv R, Fiat A, Karp RM, Turpin G (2001) Optimal search and one-way trading online algorithms. Algorithmica 30(1):101–139.
Gabrel V, Murat C, Thiele A (2014) Recent advances in robust optimization: An overview. *European journal of operational research* 235(3):471–483.

Gaspars-Wieloch H (2014) Modifications of the hurwicz’s decision rule. *Central European Journal of Operations Research* 22(4):779–794.

Hurwicz L (1951) The generalized bayes minimax principle: a criterion for decision making under uncertainty. *Cowles Commission Discussion Paper, Statistics 335*.

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

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

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

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

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

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

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

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

Roy B (2010) Robustness in operational research and decision aiding: A multi-faceted issue. *European Journal of Operational Research* 200(3):629–638.

Savage LJ (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.

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

Soyster AL (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.

Wang W, 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, Zhang JX (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 BL, den Hertog D (2019) A survey of adjustable robust optimization. *European Journal of Operational Research* 277(3):799–813.

Yue J, Chen B, Wang MC (2006) Expected value of distribution information for the newsvendor problem. *Operations Research* 54(6):1128–1136.

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