Robust Stochastic Optimization
The Synergy of Robust Optimization and Stochastic Programming

Zhi Chen
Imperial College Business School, Imperial College London
zhi.chen@imperial.ac.uk

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

We present a new mathematical optimization model known as the robust stochastic optimization (RSO), which unifies both stochastic programming and robust optimization in a synergistic and tractable framework that can be solved using the state-of-the-art commercial optimization solvers. The model of uncertainty incorporates both discrete and continuous random variables, typically assumed in stochastic programming and robust optimization respectively. To address the non-anticipativity of recourse decisions, we introduce the event-wise recourse adaptations, which integrate the scenario-tree adaptation originating from stochastic programming and the affine adaptation popularized in robust optimization. Our proposed event-wise ambiguity set is rich enough to capture traditional statistic-based ambiguity sets such as convex generalized moments, mixture distribution, φ-divergence, Wasserstein (Kantorovich-Rubinstein) metric, and also opens up novel machine-learning-based ones such as K-means clustering. We also provide several interesting models that are unique in our framework including optimizing over the Hurwicz criterion and portfolio management with adaptive rebalancing. We develop a new algebraic modeling package, RSOME to facilitate implementation of RSO models.

Key words: stochastic programming, robust optimization, machine learning.

History: February 18, 2019.

1. Introduction
In the era of modern data analytics, the ubiquity of general purpose deterministic mathematical optimization frameworks such as linear, mixed-integer and conic optimization models as well as their impact on improving management decision making cannot be understated. Algebraic modeling packages and state-of-the-art optimization solvers have been developed on these successful frameworks to facilitate implementations of prescriptive analytics to address a variety of real-world problems. Comparatively, frameworks to support generic modeling and optimization under uncertainty, despite their importance, are relatively less established. These frameworks include stochastic programming, robust optimization and more recently, distributionally robust optimization, each of them has its strengths and weaknesses.
Stochastic programming extends the linear optimization framework to minimize the total average cost associated with the optimal here-and-now and wait-and-see (or recourse) decisions under a discrete probability distribution (Danzig 1955). For enormous or infinite number of scenarios, we can obtain approximate here-and-now solutions using the sample average approximation (SAA) (see, for example, Kall and Wallace 1994, Birge and Louveaux 2011, Shapiro and Homem-de Mello 1998, Kleywegt et al. 2002). These approximate solutions are necessarily random and optimistically biased, i.e., the actual realized objectives are statistically worse off than those attained by using SAA. Stochastic programming has the versatility of modeling different types of recourse decisions including those with discrete outcomes, albeit at the expense of greater computational effort.

In classical robust optimization (Soyster 1973, Ben-Tal and Nemirovski 1998, El Ghaoui et al. 1998, Bertsimas and Sim 2004, Ben-Tal et al. 2015), the solution is obtained by reformulating the model to a deterministic optimization problem that can be solved using available solvers. The underlying uncertainty is a distribution-free continuous random variable with support confined within a convex uncertainty set. The solution obtained via classical robust optimization hedges against the worst-case outcome within the uncertainty set and hence is pessimistically biased, i.e., the realized objective value by the robust solution would often be better than the objective value attained by solving the robust optimization problem. To reduce the conservativeness, distributionally robust optimization incorporates an ambiguity set of probability distributions and its solution hedges against the worst-case distribution within the ambiguity set (Dupačová 1976, Shapiro and Kleywegt 2002, El Ghaoui et al. 2003, Delage and Ye 2010, Wiesemann et al. 2014). Under an embedded linear optimization framework, both robust optimization and distributionally robust optimization have been extended to address problems with recourse decisions (see, for instance, Ben-Tal et al. 2004, Bertsimas et al. 2018). As these models are generally computationally intractable (Shapiro and Nemirovski 2005, Ben-Tal et al. 2004), approximate solutions are sought by restricting the recourse decisions to affine mappings of the uncertainty and by requiring the recourse decisions to remain feasible almost surely. However, to ensure a tractable reformulation, these adaptive robust optimization models are also restricted to having fixed recourse and are unable to handle nonlinear recourse.

Our goal in this paper is to introduce a new mathematical optimization model known as the robust stochastic optimization (RSO), which unifies both stochastic programming and robust optimization in a synergistic and tractable framework that can be solved using currently available commercial solvers. The RSO framework incorporates both discrete and continuous random variables and introduces the event-wise static and event-wise affine adaptations to address the non-anticipativity of recourse decisions. We propose the event-wise ambiguity set that is rich enough to capture traditional statistic-based ambiguity sets and that also opens up novel machine-learning-based
ones such as K-means clustering. We showcase several interesting models that are unique in our framework. We develop the RSO framework with the mindset that it can be integrated in a general purpose software that would be accessible to modelers. As a proof of concept, we develop a new algebraic modeling package, RSOME (Robust Stochastic Optimization Modeling Environment) to facilitate modeling of problems under the RSO framework.

**Notations.** We use boldface uppercase and lowercase characters to denote matrices and vectors, respectively. We denote by $[N] \triangleq \{1, 2, \ldots, N\}$ the set of positive running indices up to $N$. We use $\mathcal{P}_0(\mathbb{R}^I)$ to represent the set of all probability distributions on $\mathbb{R}^I$. A random variable, $\tilde{z}$ is denoted with a tilde sign and we use $\tilde{z} \sim P, P \in \mathcal{P}_0(\mathbb{R}^I)$ to define $\tilde{z}$ as an $I_z$ dimensional random variable with distribution $P$.

### 2. Framework for Robust Stochastic Optimization

We now introduce the **robust stochastic optimization** (RSO) model, which combines both stochastic programming and robust optimization in a unified framework. The uncertainty associated with the RSO model comprises both discrete and continuous random variables. Specifically, $\tilde{s}$ represents a discrete random scenario taking values in $[S]$, while $\tilde{z}$ represents a continuous random variable with outcomes in $\mathbb{R}^{I_z}$. Conditioning on the realization of a scenario, $s \in [S]$, the support set of the random variable $\tilde{z}$ is tractable conic representable and is denoted by $Z_s$. The joint distribution of $(\tilde{z}, \tilde{s})$ is denoted by $P \in \mathcal{F}$, where $\mathcal{F}$ is the *ambiguity set* of probability distributions that share some identical distributional information. As espoused by Wiesemann et al. (2014) and Bertsimas et al. (2018), the random variable $\tilde{z} \triangleq (\tilde{u}, \tilde{v})$ includes both the primary $I_u$ dimensional random variable, $\tilde{u}$, and the auxiliary (or lifted) $I_v$ dimensional random variable $\tilde{v}$ associated with $\tilde{u}$. As in the same spirit of linear optimization models, the provision of the auxiliary random variable $\tilde{v}$ would greatly enhance the modeling power of the RSO model.

We denote by $w \in \mathbb{R}^J_w$ the here-and-now decision of the RSO model. The recourse decisions depend on the realization of the random variables $\tilde{z}$ and $\tilde{s}$. As it would become clearer, to obtain a tractable reformulation, we introduce two types of recourse decisions, which are function maps respectively denoted by $x(s) : [S] \to \mathbb{R}^{I_x}$ and $y(s, z) : [S] \times \mathbb{R}^{I_z} \to \mathbb{R}^{J_y}$. Here, $x(\cdot)$ adapts only to the outcome of the random scenario $\tilde{s}$, while $y(\cdot, \cdot)$ adapts to the outcomes of $\tilde{s}$ and $\tilde{z}$. Similar to most tractable adaptive robust optimization problems, the RSO model requires that for each given scenario $s \in [S]$, the function map $y(s, z)$ is affinely dependent on $z$ as follows:

$$y(s, z) \triangleq y^0(s) + \sum_{i \in [I_z]} y_i^*(s)z_i,$$

where $y^0(s), \ldots, y_i^*(s)$ account for the raw decision variables associated with $y(\cdot, \cdot)$ at scenario $s$. 
To characterize the objective function and constraints, we first define the following random variable mappings for all \( m \in [M] \cup \{0\} \),

\[
\begin{align*}
\alpha_m(s, z) &\triangleq a_{0m}^0 + \sum_{i \in [I_z]} a_{im}^i z_i \\
\beta_m(s, z) &\triangleq b_{0m}^0 + \sum_{i \in [I_z]} b_{im}^i z_i \\
\gamma_m(s) &\triangleq c_{ms}^0 \\
\delta_m(s, z) &\triangleq d_{0m}^0 + \sum_{i \in [I_z]} d_{im}^i z_i
\end{align*}
\]

(1)

for given parameters \( a_{im}^i \in \mathbb{R}^{J_w}, b_{im}^i \in \mathbb{R}^{J_x}, c_{ms}^0 \in \mathbb{R}^{J_y}, d_{im}^i \in \mathbb{R} \) \( \forall i \in [I_z] \cup \{0\}, s \in [S] \).

The objective function of the RSO model to be minimized,

\[
\sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ \left( a_{0}^\top(s, \tilde{z})w + b_{0}^\top(s, \tilde{z})x(s) + c_{0}^\top(s, \tilde{z})y(s, \tilde{z}) + d_{0}(s, \tilde{z}) \right) \right],
\]

reflects the ambiguity aversion of the decision maker against an ambiguity set \( \mathcal{F} \) that we will introduce subsequently.

There are two types constraints: hard and soft ones, which are respectively associated with the partition of indices \( M_1 \subseteq [M] \) and \( M_2 \subseteq [M] \).

The hard constraints of the RSO model, which must be satisfied almost surely, is given by the following set of semi-infinite constraints for all \( m \in M_1 \),

\[
\alpha_m^\top(s, z)w + \beta_m^\top(s, z)x(s) + \gamma_m^\top(s, z) + d_m(s, z) \leq 0 \quad \forall z \in Z_s, s \in [S].
\]

Observe that for each given scenario, the semi-infinite constraint corresponds to the standard linear robust counterpart, which can be transformed to a modest sized constraint system that can be handled by modern solvers. Stochastic programming is a special case of the RSO model in the absence of the random variable \( \tilde{z} \) and the recourse decision \( y(\cdot, \cdot) \). Likewise, adaptive robust optimization is also a special case for which \( S = 1 \).

To enhance the modeling, RSO also supports soft constraints, which must be satisfied in expectation over all distributions within the ambiguity set \( \mathcal{F} \):

\[
\sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ \left( a_{m}^\top(s, \tilde{z})w + b_{m}^\top(s, \tilde{z})x(s) + c_{m}^\top(s, \tilde{z})y(s, \tilde{z}) + d_m(s, \tilde{z}) \right) \right] \leq 0 \quad \forall m \in M_2.
\]

As in the objective function, soft constraints are evaluated in the expected sense, and hence, they capture the risk neutrality of the decision maker under ambiguity aversion. By introducing additional recourse decisions that are also embedded in the hard constraints, the RSO model is capable to capture risk-averse objective functions or safeguarding constraints; see Section 5.
Figure 1 Scenario tree. The recourse decision $x_1(\cdot)$ satisfies $x_1(1) = x_1(2) = x_1(3)$ and $x_1(4) = x_1(5)$, while the recourse decision $x_2(\cdot)$ can take differently depending on the outcome from scenarios 1, 2, ..., 5. The second level of the scenario tree gives a collection of two MECE events $\{1, 2, 3\}$ and $\{4, 5\}$, while the third level gives a collection of five MECE events $\{1\}$, $\{2\}$, $\{3\}$, $\{4\}$, and $\{5\}$.

Apart from hard and soft constraints, for a given scenario $s$, we can impose additional constraints jointly on $w$, $x(s)$ and $y^0(s), \ldots, y^I(s)$. Specifically, we have

$$r(s) \triangleq (w, x(s), y^0(s), \ldots, y^I(s)) \in X_s \quad \forall s \in [S],$$

where the feasible set $X_s$ may encompass nonlinear constraints such as conic and integral ones.

### 3. Event-Wise Recourse Adaptations

Stochastic programming and robust optimization have different approaches for addressing dynamic decision making where uncertainty is revealed in stages, and accordingly, the recourse decisions should be non-anticipative. In robust optimization, this can be achieved by restricting the dependency of a recourse decision on only a subset of the uncertainty $\tilde{z}$ that has been revealed. In stark contrast, dynamic modeling in stochastic programming is more involved and requires enumerating the complete sample paths from the beginning to the end of the decision horizon. In this regard, a scenario represents a sample path, and a scenario tree is typically used to showcase sample paths as well as decisions (Høyland and Wallace 2001, Pflug 2001, Heitsch and Römisch 2009).

Figure 1 presents the scenario tree for a three-stage problem with five scenarios: in accordance of non-anticipativity, the first-stage decision $w$ is independent of the scenarios; while the second-stage decision $x_1(\cdot)$ shall be indifferent among scenarios 1, 2, and 3 and be indifferent between scenarios 4 and 5, and the third stage decision $x_2(\cdot)$ can adapt to scenarios 1, 2, ..., 5.

To formally specify the event-wise adaptation of the recourse decision, $x(\cdot)$, we first define an event $E \subseteq [S]$ by a subset of scenarios. A partition of scenarios then induces a collection $C$ of mutually exclusive and collectively exhaustive (MECE) events. Correspondingly, we define a mapping $H_C : [S] \rightarrow C$ such that $H_C(s) = E$, for which $E$ is the only event in $C$ that contains the scenario $s$. Given a collection $C$ of MECE events, we define the event-wise static adaptation,

$$A(C) \triangleq \left\{ x : [S] \rightarrow \mathbb{R} \mid x(s) = x^E, \ E = H_C(s) \right\} \text{ for some } x^E \in \mathbb{R}.$$
Similarly, for the recourse decision, $y(\cdot, \cdot)$, we define the event-wise affine adaptation,

$$
\tilde{A} (C, I) \triangleq \left\{ y : [S] \times \mathbb{R}^{I_s} \mapsto \mathbb{R} \mid \begin{align*}
y(s, z) &= y^0(s) + \sum_{i \in I} y^i(s)z_i, \\
&\text{for some } y^0, y^i \in A(C), i \in I
\end{align*} \right\}
$$

for a subset $I \subseteq [I_s]$. Such an information index set $I$ tracks the indices of revealed uncertainties when deciding the recourse decision $y(\cdot, \cdot)$ and captures the non-anticipativity. Note that each level of the scenario tree naturally gives a partition of scenarios that induces a collection of MECE events, which in turn is used in the input of the event-wise recourse adaptation for recourse decisions associated with that level (i.e., stage); see Figure 1.

Armed with the event-wise recourse adaptations, we propose the following RSO framework:

$$
\begin{align*}
\min & \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ a_0^\top (\tilde{s}, \tilde{z}) w + b_1^\top (\tilde{s}, \tilde{z}) x(\tilde{s}) + c_j^\top (\tilde{s}) y(\tilde{s}, \tilde{z}) + d_0(\tilde{s}, \tilde{z}) \right] \\
\text{s.t.} & \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ a_m^\top (s, z) w + b_1^\top (s, z) x(s) + c_j^\top (s) y(s, z) + d_m(s, z) \leq 0 \right] \quad \forall z \in Z_s, \ s \in [S], \ m \in M_1 \\
& \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ a_m^\top (\tilde{s}, \tilde{z}) w + b_1^\top (\tilde{s}, \tilde{z}) x(\tilde{s}) + c_j^\top (\tilde{s}) y(\tilde{s}, \tilde{z}) + d_m(\tilde{s}, \tilde{z}) \leq 0 \right] \quad \forall m \in M_2 \\
& (w, x(s), y^0(s), \ldots, y^j(s)) \in \mathcal{X}_s \\
x_j \in A(C_j) \\
y_j \in \tilde{A}(C'_j, I'_j) \\
\forall s \in [S], & \forall j \in [J_x], & \forall j \in [J_y], \\
\end{align*}
$$

for given $C'_j, j \in [J_x]$ and $C'_j, j \in [J_y]$ of MECE events, and information index sets $I'_j, j \in [J_y]$.

For a given scenario $s$, the objective function and constraints are bi-affine functions of the decision variable $r(s) \in \mathbb{R}^{I_r}$. That is to say, we can write

$$
\begin{align*}
a_m^\top (s, z) w + b_1^\top (s, z) x(s) + c_j^\top (s) y(s, z) + d_m(s, z) &\triangleq r^\top (s) G_m(s) z + h_m(s) \quad \forall m \in [M] \cup \{0\},
\end{align*}
$$

for parameters $G_m(s) \in \mathbb{R}^{I_x \times I_s}$ and $h_m(s) \in \mathbb{R}$. This relation would enable us to reformulate the hard constraints into deterministic constraint systems using standard robust optimization techniques.

The RSO model is expansive, immensely versatile, and can be reformulated as a deterministic optimization problem using our developed algebraic modeling toolbox, RSOME. Before we could do so, we will next introduce the ambiguity set for the objective function and constraints.

### 4. Event-Wise Ambiguity Set

We propose the event-wise ambiguity set, which is representable in the format

$$
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0 (\mathbb{R}^{I_s} \times [S]) \mid \begin{align*}
(\tilde{z}, \tilde{s}) &\sim \mathbb{P} \\
\mathbb{E}_P \left[ \tilde{z} \mid \tilde{s} \in E_k \right] &\in Q_k \quad \forall k \in [K] \\
\mathbb{P} \left[ \tilde{z} \in Z_s \mid \tilde{s} = s \right] &= 1 \quad \forall s \in [S] \\
\mathbb{P} \left[ \tilde{s} = s \right] &= p_s \quad \forall s \in [S] \\
&\text{for some } \mathbb{P} \in \mathcal{P}
\end{align*} \right\}
$$
for given events $\mathcal{E}_k, k \in [K]$ and given closed and convex sets $\mathcal{Z}_s, s \in [S], \mathcal{Q}_k, k \in [K]$, and $\mathcal{P} \subseteq \{ p \in \mathbb{R}^S_+ \mid \sum_{s \in [S]} p_s = 1 \}$. We can effectively determine the worst-case expectation over the event-wise ambiguity set $\mathcal{F}$ by solving a classical robust optimization problem.

**Theorem 1.** Assuming the Slater’s condition holds, the worst-case expectation

$$\sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ \mathbf{r}^\top(\bar{s}) \mathbf{G}_m(\bar{s})\bar{z} + h_m(\bar{s}) \right]$$

is equivalent to the optimal value of the following classical robust optimization problem:

$$\inf \gamma \quad \text{s.t.} \quad \gamma \geq \alpha^\top p + \sum_{k \in [K]} \beta_k^\top \mu_k \quad \forall p \in \mathcal{P}, \quad \frac{\mu_k}{\sum_{s \in \mathcal{E}_k} p_s} \in \mathcal{Q}_k, \quad k \in [K]$$

$$\alpha_s + \sum_{k \in \mathcal{K}_s} \beta_k^s z \geq \mathbf{r}^\top(\bar{s}) \mathbf{G}_m(\bar{s})z + h_m(\bar{s}) \quad \forall z \in \mathcal{Z}_s, \quad s \in [S]$$

$$\gamma \in \mathbb{R}, \quad \alpha \in \mathbb{R}^S, \quad \beta_k \in \mathbb{R}^{I_s} \quad \forall k \in [K],$$

where for each $s \in [S], \mathcal{K}_s = \{ k \in [K] \mid s \in \mathcal{E}_k \}$.

The event-wise ambiguity set does not sacrifice any modeling flexibility for computational tractability. Indeed, the event-wise ambiguity set includes a wide spectrum of existing ambiguity sets in its intuitive expression.

**Uncertain Discrete Distribution**

The event-wise ambiguity set can naturally specify uncertain discrete distributions as follows:

$$\mathcal{F} = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_s} \times [S] \right) \left| \begin{array}{l} (\bar{z}, \bar{s}) \sim P \\
\mathbb{P} [\bar{z} \in \mathcal{Z}_s \mid \bar{s} = s] = 1 \quad \forall s \in [S] \\
\mathbb{P} [\bar{s} = s] = p_s \quad \forall s \in [S] \\
\text{for some } p \in \mathcal{P} \end{array} \right. \right\},$$

where each $\mathcal{Z}_s = \{ \bar{z}_s \}$ is a singleton set. As proposed in Ben-Tal et al. (2013), we can use $\phi$-divergence to characterize the uncertainty set $\mathcal{P}$ of discrete probability distributions.

**Generalized Moments Ambiguity Set**

Given a convex function $\phi : \mathbb{R}^{I_u} \rightarrow \mathbb{R}^{I_v}$, the ambiguity set based on generalized moments

$$\mathcal{G} = \left\{ P \in \mathcal{P}_0 (\mathbb{R}^{I_u}) \left| \begin{array}{l} \bar{u} \sim \mathbb{P} \\
\mathbb{E}_P [\bar{u}] \in \mathcal{Q} \\
\mathbb{E}_P [\phi(\bar{u})] \leq \sigma \\
\mathbb{P} [\bar{u} \in \mathcal{U}] = 1 \end{array} \right. \right\}.$$
can be mapped into the following event-wise ambiguity set with \( S = 1 \).

\[
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0 \left( \mathbb{R}^{I_u + I_v} \times \{1\} \right) \middle| \begin{array}{l}
(\tilde{u}, \tilde{v}, \tilde{s}) \sim \mathbb{P} \\
\mathbb{E}_\mathbb{P}[\tilde{u} | \tilde{s} = 1] \in \mathcal{Q} \\
\mathbb{E}_\mathbb{P}[\tilde{v} | \tilde{s} = 1] \leq \sigma \\
\mathbb{P}[(\tilde{u}, \tilde{v}) \in Z | \tilde{s} = 1] = 1 \\
\mathbb{P}[^{\exists} \tilde{s} = 1] = 1
\end{array} \right\},
\]

where \( Z = \{(\tilde{u}, \tilde{v}) | u \in \mathcal{U}, v \geq \phi(u)\} \). Based on the lifting and projection theorem by Wiesemann et al. (2014) (see Theorem 5 therein), we have \( \Pi_u \mathcal{F} = \mathcal{G} \). The convex generalized moments via the function \( \phi \) can provide interesting and useful statistical characterizations of the uncertainty \( \tilde{u} \), including variance, expected absolute deviation, semi-variance, and expected utility, among others.

### Wasserstein Ambiguity Set

We consider a data-driven setting as in Mohajerin Esfahani and Kuhn (2018) on the design of a Wasserstein ambiguity set centered around the empirical distribution \( \hat{\mathbb{P}} \). Given a tractable distance metric \( \rho : \mathbb{R}^{I_u} \times \mathbb{R}^{I_u} \mapsto [0, +\infty) \), the Wasserstein metric (a.k.a Kantorovich-Rubinstein metric) between any two distributions \( \mathbb{P} \) and \( \hat{\mathbb{P}} \) is defined via an optimization problem:

\[
d_W(\mathbb{P}, \hat{\mathbb{P}}) \triangleq \inf_{Q \in \mathcal{Q}(\mathbb{P}, \hat{\mathbb{P}})} \mathbb{E}_Q \left[ \rho(\tilde{u}, \tilde{u}^1) \right],
\]

where \( \tilde{u} \sim \mathbb{P}, \tilde{u}^1 \sim \hat{\mathbb{P}} \), and \( \mathcal{Q}(\mathbb{P}, \hat{\mathbb{P}}) \) is the set of all joint probability distributions on \( \mathbb{R}^{I_u} \times \mathbb{R}^{I_u} \) with marginals \( \mathbb{P} \) and \( \hat{\mathbb{P}} \). The Wasserstein ambiguity set is then defined by

\[
\mathcal{G}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0(\mathcal{U}) \middle| \begin{array}{l}
\tilde{u} \sim \mathbb{P} \\
d_W(\mathbb{P}, \hat{\mathbb{P}}) \leq \theta
\end{array} \right\},
\]

which is a ball of radius \( \theta \geq 0 \) around \( \hat{\mathbb{P}} \). Interestingly, we can provide a new representation of the Wasserstein ambiguity set in the format of an event-wise ambiguity set.

\[
\mathcal{F}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0 \left( \mathbb{R}^{I_u + 1} \times [S] \right) \middle| \begin{array}{l}
((\tilde{u}, \tilde{v}, \tilde{s}) \sim \mathbb{P} \\
\mathbb{E}_\mathbb{P}[\tilde{v} | \tilde{s} \in [S]] \leq \theta \\
\mathbb{P}[(\tilde{u}, \tilde{v}) \in Z_s | \tilde{s} = s] = 1 \ \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] = \frac{1}{S} \ \forall s \in [S]
\end{array} \right\},
\]

where the primary random variable \( \tilde{u} \) and the auxiliary random variable \( \tilde{v} \) jointly reside in lifted support sets \( Z_s = \{(u, v) | u \in \mathcal{U}, v \geq \rho(u, \tilde{u}_s)\}, s \in [S] \) for different scenarios.

**Theorem 2.** The Wasserstein ambiguity set \( \mathcal{G}_W(\theta) \) is equivalent to the marginal distribution of \( \tilde{u} \) under \( \mathbb{P} \), for all \( \mathbb{P} \in \mathcal{F}_W(\theta) \). That is, \( \mathcal{G}_W(\theta) = \Pi_u \mathcal{F}_W(\theta) \) for all \( \theta \geq 0 \).
Mixture Distribution Ambiguity Set

We can use the event-wise ambiguity set to specify a mixture distribution as proposed in Hanasusanto et al. (2015), which is useful, for example, in modeling ambiguous multi-modal distributions.

\[
F = \left\{ P \in \mathcal{P}_0(\mathbb{R}^{I_u+I_v} \times [S]) \mid \begin{align*}
(\hat{u}, \hat{v}), \tilde{s} &\sim \mathbb{P} \\
\mathbb{E}[\hat{u} | \tilde{s} = s] &\in Q_s \quad \forall s \in [S] \\
\mathbb{E}[\hat{v} | \tilde{s} = s] &\leq \sigma_s \quad \forall s \in [S] \\
P[(\hat{u}, \hat{v}) \in Z_s | \tilde{s} = s] &= 1 \quad \forall s \in [S] \\
P[\tilde{s} = s] &= p_s \quad \forall s \in [S]
\end{align*} \right\},
\]

where for each \( s \in [S] \), \( Z_s = \{ (\hat{u}, \hat{v}) \mid u \in \mathcal{U}_s, v \geq \phi(u) \} \). Note that any distribution \( P \in F \) can be written as \( P = \sum_{s \in [S]} p_s P_s \), where each mixture component \( P_s \) is an ambiguous distribution with support \( \mathcal{U}_s \) and moments \( \mathbb{E}[\hat{u} | \tilde{s} = s] \in Q_s \) and \( \mathbb{E}[\hat{v} | \tilde{s} = s] \leq \sigma_s \). Hanasusanto et al. (2015) have used the mixture distribution ambiguity set to model the uncertain demand in the textile apparel industry, which is known for the multimodality and ambiguity. As another example in a vehicle allocation problem (see a recent paper of Hao et al. 2019), the random scenario \( \tilde{s} \) indicates the random weather condition that could be sunny, rainy, and snowy, and the random variable \( \hat{u} \) denotes the uncertain taxi demand that may vary accordingly under different weather conditions.

K-means Ambiguity Set

We can naturally incorporate clustering techniques in machine learning to construct event-wise ambiguity sets directly from data. Given \( N \) historical observations \( \hat{u}_1, \ldots, \hat{u}_N \) for the primary random variable \( \hat{u} \), we can partition the support set \( \mathcal{U} \) into \( S \) clusters \( \mathcal{U}_s, s \in [S] \) using the K-means clustering (MacQueen et al. 1967, Dubes and Jain 1988), which gives the centroids \( \hat{\mu}_s, s \in [S] \) of clusters. Associated with each cluster, we can determine the support set

\[
\mathcal{U}_s = \{ u \in \mathcal{U} \mid 2u^\top (\hat{\mu}_r - \hat{\mu}_s) \leq \hat{\mu}_r^\top \hat{\mu}_r - \hat{\mu}_s^\top \hat{\mu}_s \quad \forall r \in [S] \},
\]

the weight of the cluster

\[
\hat{p}_s = \frac{1}{N} \sum_{n \in [N]} 1(\hat{u}_n \in \mathcal{U}_s),
\]

where \( 1 \) denotes the indicator function, and its convex generalized moments

\[
\hat{\sigma}_s = \frac{1}{\hat{p}_s N} \sum_{n \in [N]} 1(\hat{u}_n \in \mathcal{U}_s) \phi(\hat{u}_n).
\]

The corresponding K-means ambiguity set is a special mixture distribution ambiguity set with cluster-wise estimates \( Q_s = \{ \hat{\mu}_s \}, p_s = \hat{p}_s, \sigma_s = \hat{\sigma}_s, s \in [S] \). To account for uncertainty in these estimates, we can further specify uncertainty sets for them. We refer interested readers to a recent work of Perakis et al. (2018) using real data to construct the K-means ambiguity set to address a joint pricing and production problem. It is worth noting that classification techniques such as multivariate regression can also be incorporated to construct event-wise ambiguity sets.
5. Modeling Examples

The RSO framework is expansive and encompasses current stochastic programming and robust optimization models. Although it is based on expectations of bi-affine functions, it can also provide a tight characterization of the robust expectation of some classes of quadratic functions known in the literature, including the seminal works of Ben-Tal and Nemirovski (1998) and Tütüncü and Koenig (2004) (see details in Appendix B). We next provide several examples in our framework, including optimizing over the Hurwicz criterion and models which have both discrete and continuous recourse and where the uncertainty is characterized using the Wasserstein ambiguity set as well as the K-means ambiguity set—both are directly constructed from data. We show that an algebraic modeling toolbox such as RSOME could greatly facilitate the implementation without worrying about the tedious reformulation.

Hurwicz Criterion

Hurwicz (1951) is first to propose a decision criterion that articulates the tradeoff between pessimistic and optimistic objectives, which under distributional ambiguity can be formulated as

$$(1 - \varphi) \sup_{P \in \mathcal{F}} E_P [f(w, \tilde{u}, \tilde{s})] + \varphi \inf_{P \in \mathcal{F}} E_P [f(w, \tilde{u}, \tilde{s})],$$

where the cost function $f(w, u, s)$ depends on the here-and-now decision $w$, and it is typically convex in $u$ for given $w \in \mathcal{X}$ and scenario $s \in [S]$. Here $\varphi \in [0, 1]$ is the level of optimism, with $\varphi = 0$ ($\varphi = 1$) being the most pessimistic (optimistic) perception of the objective value. In order to obtain a computationally tractable model, we often consider the most pessimistic objective (i.e., $\varphi = 0$) because the best-case expectation for the most optimistic objective is typically non-convex in its decision $w$. Quite notably, there is a class of ambiguity sets for which the best-case expectation would also be tractable.

Proposition 1. Consider an event-wise ambiguity set $\mathcal{F}$ in (3) such that for any $P \in \mathcal{F}$, it satisfies $E_P [\tilde{u} | \tilde{s} = s] = \mu_s$ and $P[\tilde{s} = s] = p_s$ with known $p_s$ and $\mu_s$ for all $s \in [S]$. Then for any function $g(u, s) : \mathbb{R}^{I_u} \times [S] \mapsto \mathbb{R}$ that is convex in $u$ for a given $s \in [S]$, we have

$$\inf_{P \in \mathcal{F}} E_P [g(\tilde{u}, \tilde{s})] = \sum_{s \in [S]} p_s g(\mu_s, s).$$

Proof. The proposition follows trivially from Jensen’s inequality.}

The mixture distribution ambiguity set with singleton sets $Q_s, s \in [S]$ and the K-means ambiguity set fit in this class, for which we can optimize over the Hurwicz criterion

$$\min_{w \in \mathcal{X}} \left\{ (1 - \varphi) \sup_{P \in \mathcal{F}} E_P [f(w, \tilde{u}, \tilde{s})] + \varphi \sum_{s \in [S]} p_s f(w, \mu_s, s) \right\}$$

by formulating via the RSO framework.
Expectation of Convex and Piecewise Affine Functions

Expectation of convex and piecewise affine functions are commonly encountered in modeling risk aversion based on the utility (Gilboa and Schmeidler 1989) or risk measure including the shortfall risk measure (Föllmer and Schied 2002) and the optimized certainty equivalent (Ben-Tal and Teboulle 2007). We show that by simply introducing a recourse decision \( y(\cdot, \cdot) \), we can achieve an equivalent formulation under the RSO framework.

**Theorem 3.** The worst-case expectation

\[
\sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ \max_{\ell \in [L]} \left\{ r^\top (s) G_\ell (s) \bar{z} + h_\ell (s) \right\} \right] \tag{8}
\]

for a finite index set \([L]\), is equivalent to the following problem

\[
\min \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ y(s, \bar{z}) \right]
\]

\[
\text{s.t. } y(s, z) \geq r^\top (s) G_\ell (s) z + h_\ell (s) \quad \forall z \in Z, \ s \in [S], \ \ell \in [L]
\]

\[
y \in \bar{A}(\bar{C}, [I_z]),
\]

where the collection \( \bar{C} \equiv \{ \{ s \} \mid s \in [S] \} \) consists of singleton MECE events.

**Expected Utility with Mean-Covariance Ambiguity Set**

Consider the mean-covariance ambiguity set that commonly appears in portfolio management

\[
\mathcal{G}(\mu, \Sigma) = \left\{ P \in \mathcal{P}_0 (\mathbb{R}^{I_u}) \mid \begin{array}{l}
\hat{u} \sim P \\
\mathbb{E}_P [\hat{u}] = \mu \\
\mathbb{E}_P [(\hat{u} - \mu)(\hat{u} - \mu)^\top] = \Sigma
\end{array} \right\}.
\]

Here the \( I_u \)-dimensional random variable \( \hat{u} \) can be a scalar or a vector and refers to the random return(s) of the risky asset(s). For any utility function \( U : \mathbb{R} \rightarrow \mathbb{R} \), Popescu (2007) has shown that the robust expected utility of the random weighted sum \( w^\top \hat{u} \) satisfies:

\[
\inf_{P \in \mathcal{G}(\mu, \Sigma)} \mathbb{E}_P [U(w^\top \hat{u})] = \inf_{P \in \mathcal{G}(w^\top \mu, w^\top \Sigma w)} \mathbb{E}_P [U(\hat{u})]. \tag{10}
\]

This property enables Natarajan et al. (2010) to obtain an attractive computationally tractable second-order cone reformulation when \( U \) is concave piecewise affine, which is surprising as a direct duality approach would result in a positive semidefinite program that is much harder to solve. We next recover this result using the RSO framework to obtain the second-order cone reformulation.

**Theorem 4.** Given a concave piecewise affine utility function \( U(u) = \min_{\ell \in [L]} \{ g_\ell u + h_\ell \} \), the robust expected utility \( \inf_{P \in \mathcal{G}(\mu, \Sigma)} \mathbb{E}_P [U(w^\top \hat{u})] \) is equivalent to

\[
\max \inf_{P \in \mathcal{F}} \mathbb{E}_P [y(\bar{u}, \bar{v})]
\]

\[
\text{s.t. } y(u, v) \leq g_\ell (ru + w^\top \mu) + h_\ell \quad \forall (u, v) \in Z, \ \ell \in [L]
\]

\[
r \geq \sqrt{w^\top \Sigma w}
\]

\[
r \in \mathbb{R}, \ y \in \bar{A}(\{1\}, \{1, 2\}),
\]
where the ambiguity set

\[
\mathcal{F} = \left\{ \mathbb{P} \in \mathcal{P}_0 (\mathbb{R}^2) \mid \begin{array}{l}
(\hat{u}, \hat{v}) \sim \mathbb{P} \\
\mathbb{E}_\mathbb{P}[\hat{u}] = 0, \mathbb{E}_\mathbb{P}[\hat{v}] \leq 1 \\
\mathbb{P}[(\hat{u}, \hat{v}) \in \mathcal{Z}] = 1
\end{array} \right\}
\]

has only one scenario (i.e., \( S = 1 \)) and takes a support set \( \mathcal{Z} = \{(u, v) \in \mathbb{R}^2 \mid v \geq u^2\} \).

**Expectation of Saddle Functions**

The RSO model is primary based on a linear optimization framework, where the objective function and soft constraints are bilinear with respect to the underlying decision variable \( r(s) \) and the random variable \( \tilde{z} \). With auxiliary decisions and auxiliary random variables, we can also consider saddle functions that are convex with respect to the decision variables and concave with respect to the random variables (see, Ben-Tal et al. 2015). Observe that unlike earlier robust and distributionally robust optimization models, the random variable mappings in (1) include affine relations involving the auxiliary random variable, \( \tilde{v} \), which is embedded in \( \tilde{z} \). This generality allows us to extend the objective function and soft constraints to saddle functions under the RSO framework.

We consider a saddle function \( f(r(s), u, s) \) such that for a given scenario \( s \), it is jointly convex with respect to the decision \( r(s) \in \mathcal{X}_s \) for any fixed \( u \in \mathcal{U}_s \) and jointly concave with respect to \( u \in \mathcal{U}_s \) for any fixed \( r(s) \in \mathcal{X}_s \) as follows.

\[
f(r(s), u, s) \triangleq \xi^\top(r(s), s)\zeta(u, s) = \sum_{\ell \in [I_s]} \xi_\ell(r(s), s)\zeta_\ell(u, s). \tag{11}
\]

Here for a given scenario \( s \) and the corresponding partition of indices \( [I_s] = \mathcal{L}_{s1} \cup \mathcal{L}_{s2} \cup \mathcal{L}_{s3} \cup \mathcal{L}_{s4} \):

- \( \xi_\ell(r(s), s) \) is nonnegative and convex in \( r(s) \) and \( \zeta_\ell(u, s) \) is nonnegative and concave in \( u \), for \( \ell \in \mathcal{L}_{s1} \);
- \( \xi_\ell(r(s), s) \) is nonnegative and affine in \( r(s) \) and \( \zeta_\ell(u, s) \) is concave in \( u \), for \( \ell \in \mathcal{L}_{s2} \);
- \( \xi_\ell(r(s), s) \) is convex in \( r(s) \) and \( \zeta_\ell(u, s) \) is nonnegative and affine in \( u \), for \( \ell \in \mathcal{L}_{s3} \);
- \( \xi_\ell(r(s), s) \) is affine in \( r(s) \) and \( \zeta_\ell(u, s) \) is affine in \( u \), for \( \ell \in \mathcal{L}_{s4} \).

**Theorem 5.** The robust expectation

\[
\sup_{\mathbb{P} \in \mathcal{G}} \mathbb{E}_\mathbb{P} \left[ f(r(\tilde{s}), \tilde{u}, \tilde{s}) \right]
\]

for the ambiguity set

\[
\mathcal{G} = \left\{ \mathbb{P} \in \mathcal{P}_0 (\mathbb{R}^{I_u} \times [S]) \mid \begin{array}{l}
(\tilde{u}, \tilde{s}) \sim \mathbb{P} \\
\mathbb{E}_\mathbb{P}[\tilde{u} \mid \tilde{s} = k] = \mathbb{Q}_k, \forall k \in [K] \\
\mathbb{P}[\tilde{u} \in \mathcal{U}_s \mid \tilde{s} = s] = 1, \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] = p_s, \forall s \in [S] \\
\text{for some } \mathbb{P} \in \mathcal{P}
\end{array} \right\},
\]
is the same as
\[
\begin{align*}
&\min \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ \mathbf{r}^\top (\mathbf{s}) \mathbf{v} \right] \\
&\text{s.t. } \mathbf{\bar{r}}(s) \geq \xi_\ell (\mathbf{r}(s), s) \quad \forall \ell \in \mathcal{L}_{s_1} \cup \mathcal{L}_{s_3}, \ s \in [S] \\
&\mathbf{\bar{r}}(s) = \xi_\ell (\mathbf{r}(s), s) \quad \forall \ell \in \mathcal{L}_{s_2} \cup \mathcal{L}_{s_4}, \ s \in [S] \\
&\mathbf{\bar{r}}(s) \in \mathbb{R}^{I_\ell} \quad \forall s \in [S],
\end{align*}
\]
for the lifted event-wise ambiguity set
\[
\mathcal{F} = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_u + I_v} \times [S] \right) \right\} \\
\left\{ \begin{array}{l}
(P, \mathbf{s}) \sim \mathbb{P} \\
\mathbb{E}_P [\mathbf{\bar{u}} | \mathbf{s} = \mathbf{s}_k] \in \mathcal{Q}_k \quad \forall k \in [K] \\
\mathbb{P} ((\mathbf{\bar{u}}, \mathbf{v}) \in \mathcal{Z}_s | \mathbf{s} = s) = 1 \quad \forall s \in [S] \\
\mathbb{P} [\mathbf{s} = s] = p_s \quad \forall s \in [S]
\end{array} \right\}
\]
with lifted support sets
\[
\mathcal{Z}_s = \left\{ (\mathbf{u}, \mathbf{v}) \in \mathbb{R}^{I_u + I_v} \right\} \\
\left\{ \begin{array}{l}
\mathbf{u} \in \mathcal{U}_s \\
v_\ell \leq \zeta_\ell (\mathbf{u}, s) \quad \forall \ell \in \mathcal{L}_{s_1} \cup \mathcal{L}_{s_2} \\
v_\ell = \zeta_\ell (\mathbf{u}, s) \quad \forall \ell \in \mathcal{L}_{s_3} \cup \mathcal{L}_{s_4}
\end{array} \right\} \quad \forall s \in [S].
\]

Two-Stage Problem with Wasserstein Ambiguity Set
Optimization models based on the Wasserstein ambiguity set have recently attracted considerable interests from both stochastic programming and robust optimization communities. While most of the existing models are static, dynamic models with the Wasserstein ambiguity set are scarce due to limited solution approaches. We next demonstrate that the RSO framework provides a tractable approximation for two-stage linear optimization problems with the Wasserstein ambiguity set, which has the potential to serve the modeling of multi-stage dynamic problems.

In particular, we consider the following second-stage problem given the here-and-now decision \( \mathbf{w} \) and the realization \( \mathbf{u} \) of the underlying primary random variable \( \mathbf{\bar{u}} \).
\[
\begin{align}
f(\mathbf{w}, \mathbf{u}) &= \min \mathbf{c}_0^\top \mathbf{y} \\
&\text{s.t. } \mathbf{a}_\ell^\top (\mathbf{u}) \mathbf{w} + \mathbf{c}_\ell^\top \mathbf{y} \geq d_\ell (\mathbf{u}) \quad \forall \ell \in [L] \\
&\mathbf{y} \in \mathbb{R}^{I_v},
\end{align}
\]
where similar to the random variable mappings in (1), for each \( \ell \)th constraint, \( \mathbf{a}_\ell \) and \( d_\ell \) are affine mappings of the realization of \( \mathbf{\bar{u}} \). For any here-and-now decision \( \mathbf{w} \), we approximate its worst-case expected second-stage cost under the Wasserstein ambiguity set through
\[
\begin{align}
&\min \sup_{P \in \mathcal{F}_W (\mathbf{\theta})} \mathbb{E}_P \left[ \mathbf{c}_0^\top \mathbf{y}(\mathbf{s}, \mathbf{z}) \right] \\
&\text{s.t. } \mathbf{a}_\ell^\top (\mathbf{u}) \mathbf{w} + \mathbf{c}_\ell^\top \mathbf{y}(\mathbf{s}, \mathbf{z}) \geq d_\ell (\mathbf{u}) \quad \forall \mathbf{z} \in \mathcal{Z}_s, \ s \in [S], \ \ell \in [L] \\
&\mathbf{y}_j \in \mathcal{A}(\mathcal{C}_j, \mathcal{I}_j) \quad \forall j \in [J_v],
\end{align}
\]
where $\tilde{z} = (\tilde{u}, \tilde{v})$ and where the collections $\tilde{C}_j, j \in [J_y]$ of MECE events and information index sets $\mathcal{I}_j \subseteq [I_u + 1], j \in [J_y]$ jointly control how the recourse decision $y(\cdot, \cdot)$ adapts to $(\tilde{u}, \tilde{v})$ and $\tilde{s}$. The here-and-now decision $w$ can then be selected by minimizing the sum of the deterministic first-stage cost $c^\top w$ and the worst-case expected second-stage cost (15).

In Appendix C, we report the encouraging performance of this approach in comparison with (i) the computationally expensive exact approach and (ii) a state-of-the-art approximation scheme for two-stage linear optimization problems with the Wasserstein ambiguity set.

Portfolio Management with K-means Adaptive Rebalancing

We consider a three-period portfolio allocation and rebalancing problem to minimize the investment risk at the last period taking into account of transaction costs. At the beginning of the first period, we decide the number of shares $w_i \geq 0$ of stock $i \in [I_u]$ to invest at price $a_i$, incurring a transaction cost $b_i w_i$. The price of stock $i$ in the second period is $\tilde{a}_i^1 \equiv a_i (\tilde{u}_i^1 + 1)$, where $\tilde{u}_i^1$ is the corresponding return. Subsequently, for each stock $i$, we rebalance its shares to $x_i \geq 0$, which incurs a transaction cost $b_i |x_i - w_i|$. In the last period, the price of stock $i$ is $\tilde{a}_i^2 \equiv a_i (\tilde{u}_i^2 + 1)$, where $\tilde{u}_i^2$ is the third period return with respect to the first period price. The effective portfolio return at the last period, taking into account of the total transaction costs, amounts to

$$w^\top \tilde{a}_1 - w^\top a + x^\top \tilde{a}_2 - x^\top \tilde{a}_1 - b^\top (w + |x - w|) = w^\top A \tilde{u}_1^1 + x^\top A (\tilde{u}_2^2 - \tilde{u}_1^1) - b^\top (w + |x - w|),$$

where $A = \text{diag}(a)$ and the operator $|\cdot|$ takes the absolute value component-wise.

Ideally, the rebalancing decision $x$ should depend on the realization of $\tilde{u}_1^1$. However, this would lead to an intractable problem. Instead, we propose a K-means adaptive approach, where the recourse decision $x(\tilde{s})$ depends on the random scenario $\tilde{s}$ that is associated with the realization of $\tilde{u}_1^1$. In particular, using the available historical returns $\{(\hat{u}_n^1, \hat{u}_n^2)\}_{n \in [N]}$, we construct a two-layer K-means ambiguity set by first partitioning $\{\hat{u}_n^1\}_{n \in [N]}$ into $K_1$ clusters, each of which we then further partition into $K_2$ clusters based on a subset of $\{\hat{u}_n^2\}_{n \in [N]}$ that are affiliated with this specific first-layer cluster. As a result, we obtain a total number of $S = K_1 K_2$ scenarios, each of which corresponds to a unique cluster determined by the first and second layers; see an illustration in Figure 2. For each of the first-layer cluster $k \in [K_1]$, we denote by $\mathcal{E}_k \subseteq [S]$ as the set of scenarios associated with that cluster. Correspondingly, we define $\kappa(s) \in [K_1]$ as the specific first-layer cluster
Figure 2  Two-layer K-means clustering. There are 3 clusters for $\{\hat{u}_1^n\}_{n \in [N]}$ in the first layer, and a subset of $\{\hat{u}_2^n\}_{n \in [N]}$ affiliated with each of these clusters is further partitioned into 4 clusters in the second layer. In total, we have 12 distinctive clusters.

that the scenario $s$ affiliates with. Observe that $C = \{E_k \mid k \in [K_1]\}$ is a collection of MECE events.

In this way, we obtain the two-layer K-means ambiguity set

$$
\mathcal{F} = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{2I_u + 2I_v} \times [S] \right) \mid \begin{array}{l}
(\hat{u}^1, \hat{u}^2, \hat{v}^1, \hat{v}^2), \hat{s}) \sim P \\
\mathbb{E}_P [\hat{u}^1 | \hat{s} \in E_k] = \hat{\mu}^1_k \\
\mathbb{E}_P [\hat{v}^1 | \hat{s} \in E_k] \leq \hat{\sigma}^1_k \\
\mathbb{E}_P [\hat{u}^2 | \hat{s} = s] = \hat{\mu}^2_k \\
\mathbb{E}_P [\hat{v}^2 | \hat{s} = s] \leq \hat{\sigma}^2_s \\
P[(\hat{u}^1, \hat{u}^2, \hat{v}^1, \hat{v}^2) \in Z_s | \hat{s} = s] = 1 \\
P[\hat{s} = s] = p_s
\end{array} \right\},
$$

where for each $s \in [S]$, the cluster-wise support set is determined by

$$
Z_s = \{(u^1, u^2, v^1, v^2) \mid u^1 \in U^1_{\kappa(s)}, u^2 \in U^2_s, v^1 \geq \phi(u^1), v^2 \geq \phi(u^2)\}.
$$

The objective function evaluates the worst-case conditional value-at-risk (CVaR) of the final return at a pre-specified risk threshold $\varepsilon \in (0, 1)$.

$$
\sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ (w^\top A\hat{u}^1 + x^\top (\hat{s}) A(\hat{u}^2 - \hat{u}^1) - b^\top (w + |x(\hat{s}) - w|) - \delta)\right],
$$

which using now standard techniques, can be rewritten as

$$
\min_\delta \delta + \frac{1}{\varepsilon} \sup_{P \in \mathcal{F}} \mathbb{E}_P \left[ |x^\top (\hat{s}) A(\hat{u}^1 - \hat{u}^2) - w^\top A\hat{u}^1 + b^\top (w + |x(\hat{s}) - w|) - \delta|\right].
$$
Using Theorem 3, we formulate the RSO model for this portfolio optimization problem as follows:

$$\min \delta + \frac{1}{\varepsilon} \sup_{\tilde{P} \in \mathcal{F}} \mathbb{E}[y(\tilde{\mathbf{s}}, \tilde{\mathbf{z}})]$$

s.t. $y(s, z) \geq 0$ \hspace{1cm} $\forall z \in Z_s$, $s \in [S]$

$y(s, z) \geq x^\top(s)A(u^1 - u^2) - w^\top A u^1 + b^\top (w + \tilde{x}(s)) - \delta$ \hspace{1cm} $\forall z \in Z_s$, $s \in [S]$

$\tilde{x}(s) \geq x(s) - w$ \hspace{1.5cm} $\forall s \in [S]$

$\tilde{x}(s) \geq w - x(s)$ \hspace{1.5cm} $\forall s \in [S]$

$b^\top \tilde{x}(s) \leq \eta$ \hspace{1.5cm} $\forall s \in [S]$

$x(s) \geq 0$ \hspace{2cm} $\forall s \in [S]$ \hspace{1cm} (16)

$a^\top w = d$

$w \geq 0$

$x_i, \tilde{x}_i \in \mathcal{A}(\tilde{C})$ \hspace{1cm} $\forall i \in [I_u]$

$y \in \tilde{\mathcal{A}}(\tilde{C}, [2I_u + 2I_v])$

where $\tilde{x} \triangleq (\tilde{u}^1, \tilde{u}^2, \tilde{v}^1, \tilde{v}^2)$ and $\tilde{C} \triangleq \{ \{s\} \mid s \in [S] \}$ consists of singleton MECE events. For simplicity, we impose a limit $\eta$ on the transaction cost to prohibit over rebalancing in the second period, and $d$ is the initial allocation budget. We provide the sample code in RSOME in Appendix D to elucidate the intuitive implementation of the RSO model via an algebraic modeling language, which would have an impact on effectively testing the efficacy of a broad range of stochastic and robust optimization models under various forms of ambiguity sets.

References


Perakis, Georgia, Melvyn Sim, Qinshen Tang, Peng Xiong. 2018. Joint pricing and production: a fusion of machine learning and robust optimization. *Available at SSRN*.


A. Proofs

**Proof of Theorem 1.** Let \( \mu = (\mu_k)_{k \in [K]} \) and \( Q = \{ \mu \mid \mu_k \in Q_k \ \forall k \in [K] \} \). We can re-express

\[
\lambda^* = \sup_{p \in \mathcal{F}} \mathbb{E}_p \left[ r^\top (\tilde{s}) G_m(\tilde{s}) \tilde{z} + h_m(\tilde{s}) \right]
\]

by \( \lambda^* = \sup_{(p, \mu) \in \mathcal{P} \times Q} \lambda(p, \mu) \), where given \( (p, \mu) \in \mathcal{P} \times Q \), we define an ambiguity set

\[
\mathcal{F}(p, \mu) = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_z} \times [S] \right) \mid \begin{aligned}
\mathbb{P}[\tilde{z} \sim \mathcal{P}] &= \mu_k \ \forall k \in [K] \\
\mathbb{P}[\tilde{z} \in Z_s | \tilde{s} = s] &= 1 \ \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] &= p_s \ \forall s \in [S]
\end{aligned} \right\}
\]

and correspondingly the worst-case expectation

\[
\lambda(p, \mu) = \sup_{p \in \mathcal{F}(p, \mu)} \mathbb{E}_p \left[ r^\top (\tilde{s}) G_m(\tilde{s}) \tilde{z} + h_m(\tilde{s}) \right].
\]

Using the law of total probability, we can construct the joint distribution \( \mathbb{P} \) of \((\tilde{z}, \tilde{s})\) from the marginal distribution \( \hat{\mathbb{P}} \) of \( \tilde{s} \) supported on \([S]\) and the conditional distributions \( \mathbb{P}_s \) of \( \tilde{z} \) given \( \tilde{s} = s \), \( s \in [S] \). In this way, we can reformulate \( \lambda(p, \mu) \) as

\[
\lambda(p, \mu) = \sup p_k \sum_{s \in [S]} p_s \mathbb{E}_{\mathbb{P}_s} \left[ r^\top (\tilde{s}) G_m(\tilde{s}) \tilde{z} + h_m(\tilde{s}) \right]
\]

s.t. \( p_s \sum_{k \in K_s} \beta_k \tilde{z} = q_k \mu_k \ \forall k \in [K] \)

\[
\mathbb{P}_s [\tilde{z} \in Z_s] = 1 \ \forall s \in [S]
\]

with \( q_k = \sum_{s \in K_s} p_s, \ k \in [K] \). We can express the dual of \( \lambda(p, \mu) \) as

\[
\lambda_1(p, \mu) = \inf_{\alpha, q_k} \sum_{s \in [S]} \alpha_s + \sum_{k \in [K]} q_k \beta_k^\top \mu_k
\]

s.t. \( \alpha_s + \sum_{k \in K_s} \beta_k^\top \tilde{z} \geq p_s (r^\top (s) G_m(s) \tilde{z} + h_m(s)) \ \forall \tilde{z} \in Z_s, \ s \in [S] \)

\[
\alpha \in \mathbb{R}^S, \ \beta_k \in \mathbb{R}^{I_z} \ \forall k \in [K]
\]

\[
= \inf \alpha^\top p + \sum_{k \in [K]} q_k \beta_k^\top \mu_k
\]

s.t. \( \alpha_s + \sum_{k \in K_s} \beta_k^\top \tilde{z} \geq r^\top (s) G_m(s) \tilde{z} + h_m(s) \ \forall \tilde{z} \in Z_s, \ s \in [S] \)

\[
\alpha \in \mathbb{R}^S, \ \beta_k \in \mathbb{R}^{I_z} \ \forall k \in [K],
\]

where the second equality follows from for all \( s \in [S] \), first changing variable from \( \alpha_s \) to \( p_s \alpha_s \) and then dividing both sides of the constraint by \( p_s \), which is allowed since \( p \in \mathcal{P} \) is strictly positive.

By weak duality, \( \lambda(p, \mu) \leq \lambda_1(p, \mu) \). By the general min-max theorem, we further observe that

\[
\lambda^*_1 = \sup_{(p, \mu) \in \mathcal{P} \times Q} \lambda_1(p, \mu) \leq \lambda^*_2,
\]
Observe that

\[ Q \]

\[ q \]

variables in problem (17) from

\[ \text{Since} \]

\[ \text{realization of } \tilde{\theta} \text{ such that } \Pi \]

\[ G \]

\[ (\tilde{\theta}) = 0 \text{ (and hence } \tilde{\theta} \text{)} \]

\[ G \]

\[ \bar{W}_{\tilde{\theta}} \]

\[ \theta \]

\[ \text{Due to the presence of products of uncertain variables (e.g., } q_k \mu_k \text{), problem (17) is nonconvex.} \]

\[ \text{Since } p > 0 \text{ (and hence } q_k > 0 \text{), an equivalent convex representation can be obtained by changing variables in problem (17) from } q_k \mu_k \text{ to } \mu_k \text{ for all } k \in [K], \text{ which turns out to be problem (4).} \]

\[ \text{Assuming the conic representation of the following system} \]

\[ \begin{aligned} \sum_{s \in \mathcal{S}_k} \xi_s & \in \mathcal{Q}_k \quad \forall k \in [K] \\ \sum_{s \in \mathcal{S}_k} \tau_s & \in \mathcal{Z}_s \quad \forall s \in [S] \\ \xi_s & \in \mathcal{Z}_s \quad \forall s \in [S] \\ \tau & \in \mathcal{P} \end{aligned} \]

\[ (18) \]

\[ \text{satisfies the Slater’s condition (see Theorem 1.4.2 in Ben-Tal and Nemirovski 2001), one can establish strong duality, } i.e., \lambda^* = \lambda_1^* = \lambda_2^* \text{ and show that problem (4) is solvable (see Theorem 1 in Bertsimas et al. 2018).} \]

\[ \square \]

**Proof of Theorem 2.** We consider an ambiguity set without the auxiliary random variable \( \tilde{v} \)

\[ \tilde{G}_W(\theta) = \left\{ \mathbb{P} \in \mathcal{P}_0([\mathbb{R}^J] \times [S]) \left| \begin{array}{l} \mathbb{P} [\tilde{u}, \tilde{s}] \sim \mathcal{P} \\ \mathbb{E}_\mathcal{P} [\xi (\tilde{u}, \tilde{s})] \leq \theta \\ \mathbb{P} [\tilde{u} \in \mathcal{U} | \tilde{s} = s] = 1 \quad \forall s \in [S] \\ \mathbb{P} [\tilde{s} = s] = \frac{1}{S} \quad \forall s \in [S] \end{array} \right. \right\}. \]

\[ (19) \]

Since this ambiguity set satisfies \( \Pi_{(\tilde{u}, \tilde{s})} \mathcal{F}_W(\theta) = \tilde{G}_W(\theta) \) for all \( \theta \geq 0 \), thus it is sufficient to prove \( \Pi_u \tilde{G}_W(\theta) = \tilde{G}_W(\theta) \) for all \( \theta \geq 0 \).

To this end, we first prove \( \mathcal{G}_W(\theta) \subseteq \Pi_u \tilde{G}_W(\theta) \). Consider \( \tilde{u} \sim \mathcal{P} \) for some \( \mathbb{P} \in \mathcal{G}_W(\theta) \). By definition of the Wasserstein ambiguity set \( \mathcal{G}_W(\theta) \), there exists a joint distribution \( \mathbb{Q} \in \mathcal{P}(\mathbb{P}, \hat{\mathbb{P}}) \) of (\( \tilde{u}, \tilde{u}^t \)) such that \( \Pi_u \mathbb{Q} = \mathbb{P}, \Pi_{u^t} \mathbb{Q} = \hat{\mathbb{P}} \), and \( \mathbb{E}_\mathcal{Q} [\rho(\tilde{u}, \tilde{u}^t)] \leq \theta \). Since we can construct \( \mathbb{Q} \) from the marginal distribution \( \hat{\mathbb{P}} \) of \( \tilde{u}^t \) supported on \( \{\tilde{u}_1, \ldots, \tilde{u}_S\} \) and the conditional distributions \( \mathbb{P}_s \) of \( \tilde{u}_s \), given the realization of \( \tilde{u}^t \) is \( \tilde{u}_s \), \( s \in [S] \), we have \( \mathbb{P} (\tilde{u}, \tilde{u}^t) \sim \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s \otimes \delta_{\tilde{u}_s} \). We can then construct a distribution \( \mathbb{Q}' \in \mathcal{P}_0([\mathbb{R}^J] \times [S]) \) for the random variable \( (\tilde{u}, \tilde{s}) \sim \mathbb{Q}' \) via \( \mathbb{Q}' = \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s \otimes \delta_{\tilde{u}_s} \). Observe that \( \mathbb{Q}' \in \tilde{G}_W(\theta) \), hence \( \mathcal{G}_W(\theta) \subseteq \Pi_u \tilde{G}_W(\theta) \).
To prove $\Pi_\theta \tilde{G}_W(\theta) \subseteq G_W(\theta)$, we fix any $\mathbb{P} \in \tilde{G}_W(\theta)$ and we write its projection over $\tilde{u}$ as $\Pi_\theta \mathbb{P} = \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s$, where $\mathbb{P}_s$ is the conditional distribution of $\tilde{z}$ given the outcome of the random scenario is $s$. We can then construct a joint distribution $\mathbb{Q} = \frac{1}{S} \sum_{s \in [S]} \mathbb{P}_s \otimes \delta_a$, of $(\tilde{u}, \tilde{u}^\dagger)$ that satisfies

$$\mathbb{E}_\mathbb{Q} [\rho(\tilde{u}, \tilde{u}^\dagger)] = \frac{1}{S} \sum_{s \in [S]} \mathbb{E}_{\mathbb{P}_s} [\rho(\tilde{u}, \tilde{u}_s)] = \mathbb{E}_\mathbb{P} [\rho(\tilde{u}, \tilde{u}_s) | \tilde{s} \in [S]] \leq \theta.$$  

Hence, $\Pi_\theta \mathbb{P} \in G_W(\theta)$, which gives $\Pi_\theta \tilde{G}_W(\theta) \subseteq G_W(\theta)$ to conclude $G_W(\theta) = \Pi_\theta \tilde{G}_W(\theta)$. \hfill \qed

**Proof of Theorem 3.** Previous derivations in the proof of Theorem 1 implies that (i) the worst-case expectation (8) is equivalent to the following problem

$$\inf \gamma \quad \text{s.t.} \quad \gamma \geq \alpha^\top p + \sum_{k \in [K]} \beta_k^\top \mu_k \quad \forall p \in \mathcal{P}, \quad \frac{\mu_k}{\sum_{s \in E_k} p_s} \in \mathcal{Q}_k, \ k \in [K]$$

$$\alpha_s + \sum_{k \in K_s} \beta_k^\top z \geq y(s, z) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \ell \in [L]$$

(20) $\gamma \in \mathbb{R}, \ \alpha \in \mathbb{R}^S, \ \beta_k \in \mathbb{R}^{I_z} \quad \forall k \in [K]$;

and (ii) problem (9) is equivalent to

$$\inf \gamma \quad \text{s.t.} \quad \gamma \geq \alpha^\top p + \sum_{k \in [K]} \beta_k^\top \mu_k \quad \forall p \in \mathcal{P}, \quad \frac{\mu_k}{\sum_{s \in E_k} p_s} \in \mathcal{Q}_k, \ k \in [K]$$

$$\alpha_s + \sum_{k \in K_s} \beta_k^\top z \geq y(s, z) \quad \forall z \in \mathcal{Z}_s, \ s \in [S]$$

(21) $\gamma \in \mathbb{R}, \ \alpha \in \mathbb{R}^S, \ \beta_k \in \mathbb{R}^{I_z} \quad \forall k \in [K]$.

It is then sufficient to construct a feasible solution to problem (21) from any feasible solution to problem (20) such that the constructive solution yields the same objective as problem (20). Indeed, given a feasible solution $(\delta^\dagger, \alpha^\dagger, (\beta_k^\dagger)_{k \in [K]})$ to problem (20), we can construct a desired solution to problem (21) via:

$$\delta = \delta^\dagger, \ \alpha = \alpha^\dagger, \ \beta_k = \beta_k^\dagger \ \forall k \in [K], \ y(s, z) = \alpha_s^\dagger + \sum_{k \in K_s} (\beta_k^\dagger)^\top z \ \forall s \in [S],$$

for which the recourse decision $y(\cdot, \cdot) \in \tilde{A}(\tilde{C}, [I_z])$. \hfill \qed

**Proof of Theorem 4.** Using the result of Popescu (2007), we first show that

$$\inf_{\mathbb{P} \in \mathcal{G}(\mathbb{w}^\top \mu, \mathbb{w}^\top \Sigma \mathbb{w})} \mathbb{E}_{\mathbb{P}} [U(\tilde{u})] = \sup_r \left\{ \inf_{\mathbb{P} \in \mathcal{F}} \mathbb{E}_{\mathbb{P}} [U(r \tilde{u} + \mathbb{w}^\top \mu)] \mid r \geq \sqrt{w^\top \Sigma \mathbb{w}} \right\}.$$
By duality, we have

\[
\inf_{P \in G(w \mu, w \Sigma w)} E_P[U(\tilde{u})] = \inf_{P \in G(0, w \Sigma w)} E_P[U(\tilde{u} + w \mu)] = \sup_{\alpha, \beta_1, \beta_2, r} \left\{ \alpha + w^\top \Sigma w \beta_2 \mid \alpha + \beta_1 u + \beta_2 u^2 \leq U(u + w \mu) \quad \forall u \right\}.
\]

Note that it requires \(\beta_2 \leq 0\) for the above problem to be feasible, as otherwise the constraint would be violated for some sufficiently large \(u\). Hence, we can further rewrite this problem into

\[
\sup_{\alpha, \beta_1, \beta_2, r, \beta_2} \left\{ \alpha + \beta_2 \mid \alpha + \beta_1 u + \beta_2 u^2 \leq U(ru + w \mu) \quad \forall u \right\}.
\]

The result then follows by applying Theorem 3.

**Proof of Theorem 5.** Observe that for any feasible recourse decision \(\tilde{r}(\cdot)\) to problem (12), we have

\[
\sup_{r \in F} E_P[\tilde{r}^\top(\tilde{s}) \tilde{v}] = \sup_{P \in G} E_P[\tilde{r}^\top(\tilde{s}) \zeta(\tilde{u}, \tilde{s})].
\]

In addition, the optimal \(\tilde{r}^*(\cdot)\) to problem (12) satisfies

\[
\tilde{r}^*_\ell(s) = \xi_\ell(r(s), s) \quad \forall \ell \in [I_v], s \in [S].
\]

Therefore, our claim holds.

**B. Expectation of Quadratic Functions**

The RSO framework can be used to provide a tight characterization of the robust expectation of some quadratic functions that are known in the literature. Let \(S^I\) be the space of symmetric matrices in \(R^{I \times I}\). Given \(X, Y \in S^I\), we denote by \(X \succeq Y\) (resp., \(X \succ Y\)) to represent \(X - Y\) is positive semidefinite (resp., definite), and denote by \(X \bullet Y\) as the trace inner product of \(X, Y\). Special matrices and vectors of the appropriate dimension include \(O, I, 0\), which respectively correspond to the zero matrix, the identity matrix, and the zero vector.
## Bi-Convex-Quadratic Function

We explore the following bi-convex-quadratic function as an extension of Ben-Tal and Nemirovski (1998) to include discrete scenarios:

\[
g(r(s), u, s) = u^\top A^\top (r(s), s) A(r(s), s) u + 2u^\top b(r(s), s) + c(r(s), s),
\]

where given a scenario \(s\), \(A(r(s), s), b(r(s), s), c(r(s), s)\) are affine mappings of \(r(s)\). The event-wise ambiguity set is given by

\[
G = \left\{ P \in \mathcal{P}_0 \left( \mathbb{R}^{I_u} \times [S] \right) \left| \begin{array}{l}
(\tilde{u}, \tilde{s}) \sim P \\
\mathbb{E}_P \left[ \left( \frac{1}{u} \right) \left( \frac{1}{u} \right)^\top \bigg| \tilde{s} \in \mathcal{E}_k \right] \in \mathcal{Q}_k & \forall k \in [K] \\
\mathbb{P} \left[ \left( \frac{1}{u} \right) \left( \frac{1}{u} \right)^\top \in \mathcal{U}_s \bigg| \tilde{s} = s \right] = 1 & \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] = p_s & \forall s \in [S] \\
\text{for some } p \in \mathcal{P}
\end{array} \right. \right\}.
\]

The support set \( \mathcal{U}_s \) is general enough to capture the ubiquitous uncertainty set \( \{u | u^\top \Lambda_s u \leq 1\} \) parameterized by some \( \Lambda_s > 0 \), for which we only need to define

\[
\mathcal{U}_s = \left\{ U \in S_{I_u+1}^* \bigg| U \cdot \begin{pmatrix} -1 & 0^\top \\ 0 & \Lambda_s \end{pmatrix} \leq 0 \right\}. \tag{22}
\]

**Theorem 6.** The worst-case expectation

\[
\sup_{P \in G} \mathbb{E}_P \left[ g(r(\tilde{s}), \tilde{u}, \tilde{s}) \right] \tag{23}
\]

is bounded from above by

\[
\min_{P \in \mathcal{P}} \left\{ \mathbb{E}_P \left[ R(\tilde{s}) \cdot \tilde{Z} \right] \right\}
\]

\[
\text{s.t. } \begin{pmatrix} R(s) - \begin{pmatrix} 1 & b^\top (r(s), s) \\ b(r(s), s) & O \end{pmatrix} & \begin{pmatrix} 0 & 0^\top \\ 0 & A^\top (r(s), s) \end{pmatrix} \end{pmatrix} \succeq O & \forall s \in [S] \tag{24}
\]

where the lifted event-wise ambiguity set

\[
\mathcal{F} = \left\{ P \in \mathcal{P}_0 \left( S_{I_u+1}^* \times [S] \right) \left| \begin{array}{l}
(\tilde{Z}, \tilde{s}) \sim P \\
\mathbb{E}_P \left[ \tilde{Z} \bigg| \tilde{s} \in \mathcal{E}_k \right] \in \mathcal{Q}_k & \forall k \in [K] \\
\mathbb{P}[\tilde{Z} \in \mathcal{Z}_s \bigg| \tilde{s} = s] = 1 & \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] = p_s & \forall s \in [S] \\
\text{for some } p \in \mathcal{P}
\end{array} \right. \right\}
\]

takes lifted support sets \( \mathcal{Z}_s = \{ Z \in \mathcal{U}_s | Z \succeq O, [Z]_{1,1} = 1 \}, s \in [S] \). Moreover, the bound is tight for ellipsoidal support sets defined in (22).
Proof of Theorem 6. We note that
\[ g(r(s), u, s) = \begin{pmatrix} 1 & b(r(s), s) \\ b(r(s), s) A^T(r(s), s) A(r(s), s) \end{pmatrix} \cdot \begin{pmatrix} 1 \\ u \end{pmatrix} \begin{pmatrix} 1 \\ u \end{pmatrix}^T. \]

By Schur complement, the positive semidefinite constraints of problem (24) is equivalent to
\[ R(s) \succeq \begin{pmatrix} 1 & b(r(s), s) \\ b(r(s), s) A^T(r(s), s) A(r(s), s) \end{pmatrix}. \]

Since \( Z \in Z_s \) is positive semidefinite, an optimal \( R(s) \) would be
\[ R(s) = \begin{pmatrix} 1 & b(r(s), s) \\ b(r(s), s) A^T(r(s), s) A(r(s), s) \end{pmatrix}. \]

Observe that the ambiguity set \( \mathcal{F} \) coincides with \( \mathcal{G} \) if every support set \( Z_s \) is replaced by \( \bar{Z}_s = \{ Z \in U_s | Z \succeq O, [Z]_{1,1} = 1, \text{rank}(Z) = 1 \} \), which however, would lead to a harder problem to solve due to the rank constraint. Since \( \bar{Z}_s \subseteq Z_s \), we obtain the conservative upper bound.

We next show that the bound is tight for ellipsoidal uncertainty sets defined in (22). After using Theorem 1 to reformulate problem (23), we need to deal with the following robust counterpart.
\[ \alpha_s \geq \Phi_s \cdot \begin{pmatrix} 1 \\ u \end{pmatrix} \begin{pmatrix} 1 \\ u \end{pmatrix}^T \quad \forall u^T \Lambda_s u \leq 1 \]

for some \( \alpha_s \in \mathbb{R}, \Phi_s \in S^{I_u+1} \), which by S-lemma, is equivalent to
\[ \begin{pmatrix} \alpha_s & 0^T \\ 0 & 0 \end{pmatrix} + \delta_s \begin{pmatrix} -1 & 0^T \\ 0 & \Lambda_s \end{pmatrix} \succeq \Phi_s, \]

for some \( \delta_s \geq 0 \). On the other hand, the robust counterpart in the reformulation of problem (24)
\[ \alpha_s \geq \Phi_s \cdot Z \quad \forall Z \cdot \begin{pmatrix} -1 & 0^T \\ 0 & \Lambda_s \end{pmatrix} \leq 0, [Z]_{1,1} = 1, Z \succeq O \]
is equivalent to
\[ \begin{pmatrix} \tau_s & 0^T \\ 0 & 0 \end{pmatrix} + \delta_s \begin{pmatrix} -1 & 0^T \\ 0 & \Lambda_s \end{pmatrix} \succeq \Phi_s, \]

for some \( \tau_s \leq \alpha_s \) and \( \delta_s \geq 0 \), for which we can replace \( \tau_s \) with \( \alpha_s \) without affecting its feasibility. This establishes the desired tight bound for ellipsoidal uncertainty sets. \( \square \)
Bi-Conic-Quadratic Function

We can also extend the bi-conic-quadratic function considered in Ben-Tal and Nemirovski (1998) to include discrete scenarios as follows:

\[ h(r(s), u, s) \triangleq \| A(r(s), s) u + b(r(s), s) \|_2, \]

where given \( s, A(r(s), s), b(r(s), s) \) are affine mappings of \( r(s) \). The event-wise ambiguity set takes

\[ \mathcal{G} = \left\{ P \in \mathcal{P}_0 ([\mathbb{R}^I \times [S]]) \left| \begin{array}{l}
(\tilde{u}, \tilde{s}) \sim P \\
\mathbb{P}\left[ \frac{1}{\tilde{u}} \left( \frac{1}{\tilde{u}} \right)^\top \in U_s \right| \tilde{s} = s \right] = 1 \quad \forall s \in [S] \\
\mathbb{P}[\tilde{s} = s] = p_s \quad \forall s \in [S] \\
\text{for some } p \in \mathcal{P}
\end{array} \right. \right\}. \]

**Theorem 7.** The worst-case expectation

\[ \sup_{P \in \mathcal{G}} \mathbb{E}_P [h(\tilde{r}(\tilde{s}), \tilde{u}, \tilde{s})] \]

is bounded from above by

\[ \min \sup_{P \in \mathcal{F}} \mathbb{E}_P [x(\tilde{s})] \]

subject to

\[ x(s) \geq R(s) \cdot Z \quad \forall Z \in \mathcal{Z}_s, s \in [S] \]

\[ \left( \begin{array}{c}
R(s) \\
A^\top(r(s), s) \\
(b(r(s), s) A(r(s), s)) \\
x(s)I
\end{array} \right) \succeq O \quad \forall s \in [S] \]

\[ R(s) \in \mathcal{S}_u^{I_s + 1} \]

\[ x \in \mathcal{A}(\bar{\mathcal{C}}), \]

where \( \bar{\mathcal{C}} \triangleq \{ \{s\} | s \in [S] \} \) and where the lifted event-wise ambiguity set

\[ \mathcal{F} = \left\{ P \in \mathcal{P}_0 ([\mathbb{R}^I \times [S]]) \left| \begin{array}{l}
(\bar{Z}, \bar{s}) \sim P \\
\mathbb{P}[\bar{Z} \in \mathcal{Z}_s | \bar{s} = s] = 1 \quad \forall s \in [S] \\
\mathbb{P}[\bar{s} = s] = p_s \quad \forall s \in [S] \\
\text{for some } p \in \mathcal{P}
\end{array} \right. \right\}. \]

takes lifted support sets \( \mathcal{Z}_s = \{ Z \in U_s | Z \succeq O, [Z]_{1,1} = 1 \}, s \in [S] \). Moreover, the bound is tight for ellipsoidal uncertainty sets defined in (22).

**Proof of Theorem 7.** Since the ambiguity set does not contain any expectation constraint, we can obtain a tractable reformulation by replacing \( h(r(s), u, s) \) with a recourse variable \( x(s) \) and imposing the following constraint (see reformulation in Theorem 1):

\[ x^2(s) \geq h^2(r(s), u, s) \quad \forall \left( \begin{array}{c}
1 \\
u
\end{array} \right) \left( \begin{array}{c}
1 \\
u
\end{array} \right)^\top \in U_s, s \in [S]. \]
We next discuss how such a constraint can be specified in problem (25). Observe that
\[ h^2(r(s), u, s) = \left( \begin{pmatrix} b^\top(r(s), s) \\ A^\top(r(s), s) \end{pmatrix}^{\top} \begin{pmatrix} b(r(s), s) & A(r(s), s) \end{pmatrix} \right) \cdot \left( \begin{pmatrix} 1 \\ u \end{pmatrix}^{\top} \right). \]

By Schur complement, the positive semidefinite constraints of problem (24) is equivalent to
\[ x(s)R(s) \succeq \begin{pmatrix} b^\top(r(s), s) \\ A^\top(r(s), s) \end{pmatrix}^{\top} \begin{pmatrix} b(r(s), s) & A(r(s), s) \end{pmatrix} R(s). \]

Since \( Z \in \mathbb{Z} \) is positive semidefinite and \( x(s) \geq 0 \), an optimal \( R(s) \) would be
\[ x(s)R(s) = \begin{pmatrix} b^\top(r(s), s) \\ A^\top(r(s), s) \end{pmatrix}^{\top} \begin{pmatrix} b(r(s), s) & A(r(s), s) \end{pmatrix} R(s). \]

The rest of the proof follows similarly as in the proof of Theorem 6. \( \square \)

**Affine-Quadratic Function**
As an extension of Tütüncü and Koenig (2004), we consider a saddle function that is convex quadratic with respect to the decision variable and that is affine with respect to \( z \):
\[ g(r(s), z, s) \triangleq r^\top(s)H(s, z)r(s) + r^\top(s)G(s)z + h(s), \tag{26} \]
where given a scenario \( s \), \( H(s, z) \) is an affine mapping of \( z \) and \( \mathcal{Z} \subseteq \{ z \mid H(z, s) \succeq O \} \). Introducing auxiliary variables \( R(s) \in \mathbb{S}^{I_u} \), \( s \in [S] \) and using the Schur complement, the robust expectation sup\( \mathbb{E}_\mathcal{F}[g(r(\tilde{s}), \tilde{z}, \tilde{s})] \) is the same as
\[
\min_{\mathcal{P} \in \mathcal{F}} \sup_{s \in [S]} \mathbb{E}_{\mathcal{F}} \left[ R(\tilde{s}) \cdot H(\tilde{s}, \tilde{z}) + r^\top(\tilde{s})G(\tilde{s})\tilde{z} + h(\tilde{s}) \right] \\
\text{s.t.} \quad \begin{pmatrix} 1 & r^\top(s) \\ r(s) & R(s) \end{pmatrix} \succeq 0 \\
R(s) \in \mathbb{S}^{I_u} \quad \forall s \in [S],
\]
which falls within the RSO framework.

**C. Computational Experiments with Wasserstein Ambiguity Set**
We focus on two-stage linear optimization problems with the data-driven Wasserstein ambiguity set in the form (7), given some past observations \( \hat{u}_1, \ldots, \hat{u}_S \) of the uncertainty.

**Multi-Item Newsvendor Problem**
We consider a multi-item newsvendor problem with \( I_u \) different items. For each item \( i, i \in [I_u] \), its unit selling price and ordering cost are denoted by \( p_i \) and \( c_i \), respectively. Under a total budget \( d \), the decision maker decides the ordering quantity \( w_i \) of each item before its random demand \( \tilde{u}_i \) is
observed. Once the demand realizes, the selling quantity of each item is decided as \( \min \{ w_i, u_i \} \).

The decision maker maximizes the worst-case expected operating revenue by solving

\[
\max \inf_{p \in F_W(\theta)} \mathbb{E}_p \left[ \sum_{i \in [I_u]} p_i \min \{ w_i, \tilde{u}_i \} \right] \\
\text{s.t.} \quad c^T w = d, \ w \geq 0,
\]

which can be recast as a minimization problem,

\[
\min -p^T w + \sup_{p \in F_W(\theta)} \mathbb{E}_p \left[ \sum_{i \in [I_u]} p_i (w_i - \tilde{u}_i)^+ \right] \\
\text{s.t.} \quad c^T w = d, \ w \geq 0,
\]

in the objective function, \( \sum_{i \in [I_u]} p_i (w_i - \tilde{u}_i)^+ = \max_{\mathcal{J} \subseteq [I_u]} \sum_{j \in \mathcal{J}} p_j (w_j - u_j) \) is convex and piecewise affine involving \( 2^{I_u} \) pieces. Thus by Theorem 3, problem (27) can be exactly solved by

\[
\lambda^* = \min -p^T w + \sup_{p \in F_W(\theta)} \mathbb{E}_p [ y(\tilde{s}, \tilde{z}) ] \\
\text{s.t.} \quad y(s, z) \geq \sum_{j \in \mathcal{J}} p_j (w_j - u_j) \quad \forall z \in \mathcal{Z}_s, \ s \in [S], \ \mathcal{J} \subseteq [I_u] \\
c^T w = d, \ w \geq 0 \\
y \in \bar{\mathcal{A}}(\bar{C}, [I_u + 1]),
\]

where we introduce a recourse variable \( y(\cdot, \cdot) \) following the event-wise affine adaptation with the collection \( \bar{C} \triangleq \{ \{ s \} | s \in [S] \} \). Problem size of this exact approach however, increases exponentially in the number of items. Alternatively, we can obtain an upper bound by solving an RSO problem:

\[
\lambda = \min -p^T x + \sup_{p \in F_W(\theta)} \mathbb{E}_p [ p^T y(\tilde{s}, \tilde{z}) ] \\
\text{s.t.} \quad y(s, z) \geq 0 \quad \forall z \in \mathcal{Z}_s, \ s \in [S] \\
y(s, z) \geq w - u \quad \forall z \in \mathcal{Z}_s, \ s \in [S] \\
c^T w = d, \ w \geq 0 \\
y_i \in \bar{\mathcal{A}}(\bar{C}, \mathcal{I}) \quad \forall i \in [I_u],
\]

where we control how the recourse decision \( y(\cdot, \cdot) \) adapts to \( (\tilde{u}, \tilde{v}) \) and \( \tilde{s} \) through choosing the collection \( \bar{C} \) of MECE events and the information index set \( \mathcal{I} \).

We consider \( I_u \in \{5, 7\} \) and \( S \in \{5, 10, 20, 50\} \). The random demand belongs to a support set \( \mathcal{U} = [0, \tilde{u}] \), and we use the Euclidean norm \( \| \cdot \|_2 \) as the distance metric. In each instance, we randomly generate the upper bound \( \tilde{u} \) from a uniform distribution on \( [0, 100]^I_u \). Subsequently, past observations are randomly generated from the uniform distribution on \( [0, \tilde{u}] \). We set \( c_i = 1, i \in [I_u] \) and \( b = 50I_u \), and we generate \( p \) from a uniform distribution on \( [0, 5]^I_u \). For different choices of \( \theta \), we run 100 random instances and compare the performance of different cases of event-wise recourse adaptation against the exact approach:
Table 1 5 items: 90-th percentile optimality gaps (in %) of case 1 (left), case 2 (middle), and case 3 (right).

<table>
<thead>
<tr>
<th>$S$</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1</td>
<td>2.0</td>
<td>&lt;0.1</td>
<td>3.4</td>
<td>8.8</td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1</td>
<td>4.5</td>
<td>11.6</td>
<td>&lt;0.1</td>
<td>6.4</td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1</td>
<td>4.8</td>
<td>5.7</td>
<td>&lt;0.1</td>
<td>5.3</td>
</tr>
<tr>
<td>50</td>
<td>&lt;0.1</td>
<td>6.8</td>
<td>7.9</td>
<td>&lt;0.1</td>
<td>7.4</td>
</tr>
</tbody>
</table>

Table 2 7 items: 90-th percentile optimality gaps (in %) of case 1 (left), case 2 (middle), and case 3 (right).

<table>
<thead>
<tr>
<th>$S$</th>
<th>1</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1</td>
<td>1.3</td>
<td>8.9</td>
<td>0.1</td>
<td>2.3</td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1</td>
<td>2.5</td>
<td>6.2</td>
<td>0.1</td>
<td>3.5</td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1</td>
<td>4.8</td>
<td>6.7</td>
<td>0.1</td>
<td>5.7</td>
</tr>
<tr>
<td>50</td>
<td>0.1</td>
<td>5.5</td>
<td>6.1</td>
<td>0.1</td>
<td>5.8</td>
</tr>
</tbody>
</table>

- case 1: $y(\cdot, \cdot)$ adapts on $\bar{u}, \bar{v}, \bar{s}$, i.e., $C = \{s \mid s \in [S]\}$ and $I = [I_u + 1]$;
- case 2: $y(\cdot, \cdot)$ adapts only on $\bar{u}, \bar{v}$, i.e., $C = \{1, \ldots, S\}$ and $I = [I_u + 1]$;
- case 3: $y(\cdot, \cdot)$ adapts only on $\bar{u}$, i.e., $C = \{1, \ldots, S\}$ and $I = [I_u]$.

Case 1 corresponds to the full event-wise affine adaptation. We turn off the event-wise adaptation in case 2, while in case 3 we further deprive the recourse decision $y(\cdot, \cdot)$ of the affine adaptation on the auxiliary random variable $\bar{v}$. For each case, we consider the following relative gap between the objective value using the event-wise recourse adaptation and the exact optimal objective value:

$$\frac{\lambda^* - \lambda}{\lambda^*} \times 100\%.$$ 

Results for $I_u = 5$ items and $I_u = 7$ items are summarized in Table 1 and Table 2, respectively. Notably, with (i) the notion of event-wise adaptation and (ii) the inclusion of auxiliary random variable $\bar{v}$, the full event-wise affine adaptation can provide a high-quality approximation to the exact approach; while excluding either (i) or (ii) may lead to a more conservative approximation.

We evaluate the scalability of the full event-wise affine adaptation (case 1), the affine adaptation without event-wise dependence (case 2), and the exact approach, by comparing their computation times and limits for different pairs of problem sizes. For the exact approach, the computer runs
Table 3  Computation times of the exact approach. We report only those \((S,I_u)\) pairs for which the exact approach were solved.

<table>
<thead>
<tr>
<th>(S)</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.6</td>
<td>2.4</td>
<td>3.7</td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1</td>
<td>0.3</td>
<td>0.7</td>
<td>1.3</td>
<td>2.2</td>
<td>3.8</td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1</td>
<td>0.8</td>
<td>1.8</td>
<td>3.7</td>
<td>6.8</td>
<td>10.9</td>
</tr>
<tr>
<td>50</td>
<td>0.2</td>
<td>1.8</td>
<td>5.2</td>
<td>10.3</td>
<td>17.8</td>
<td>31.0</td>
</tr>
<tr>
<td>100</td>
<td>0.4</td>
<td>1.8</td>
<td>5.2</td>
<td>10.3</td>
<td>17.8</td>
<td>31.0</td>
</tr>
<tr>
<td>200</td>
<td>0.5</td>
<td>2.0</td>
<td>5.5</td>
<td>11.0</td>
<td>19.1</td>
<td>32.8</td>
</tr>
<tr>
<td>300</td>
<td>0.8</td>
<td>3.1</td>
<td>7.8</td>
<td>16.5</td>
<td>32.3</td>
<td>49.6</td>
</tr>
<tr>
<td>500</td>
<td>1.4</td>
<td>5.7</td>
<td>16.6</td>
<td>36.6</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>800</td>
<td>2.5</td>
<td>14.8</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
</tbody>
</table>

Table 4  Computation times and limits of the affine approximation without event-wise adaptation (left) and the full event-wise affine adaptation (right). The symbol “—” indicates “out of memory”.

<table>
<thead>
<tr>
<th>(S)</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>25</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>&lt;0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>10</td>
<td>&lt;0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>20</td>
<td>&lt;0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.6</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td>50</td>
<td>0.1</td>
<td>0.4</td>
<td>0.8</td>
<td>1.9</td>
<td>3.0</td>
<td>4.4</td>
</tr>
<tr>
<td>100</td>
<td>0.2</td>
<td>1.1</td>
<td>2.7</td>
<td>5.1</td>
<td>9.2</td>
<td>10.5</td>
</tr>
<tr>
<td>200</td>
<td>0.4</td>
<td>1.7</td>
<td>5.2</td>
<td>15.8</td>
<td>30.7</td>
<td>25.3</td>
</tr>
<tr>
<td>300</td>
<td>0.8</td>
<td>3.2</td>
<td>8.6</td>
<td>14.2</td>
<td>36.1</td>
<td>48.4</td>
</tr>
<tr>
<td>500</td>
<td>1.1</td>
<td>5.3</td>
<td>13.8</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
<tr>
<td>800</td>
<td>4.3</td>
<td>13.8</td>
<td>——</td>
<td>——</td>
<td>——</td>
<td>——</td>
</tr>
</tbody>
</table>

out of memory when the number of items exceeds 10 and the number of samples exceeds 5 (see Table 3), which is not practically favorable. In contrast, we are able to obtain a high-quality solution via the full event-wise affine adaptation with modest computational effort. Quite interestingly, the event-wise adaptation that plays the key role in delivering the high-quality approximation seems to require only a little extra computational effort (see Table 4).

The following code segment shows how to implement the full event-wise affine adaptation for the multi-item newsvendor problem with Wasserstein ambiguity sets in RSOME.

```matlab
% I: number of items
% S: number of past observations
% theta: radius
% Gamma: total budget
% cost (price): cost (price) parameters
% ubar: upper bound of demand
% U = (u_1, ..., u_S): past realizations
```
Create RSOME model

```
model = rsome('newsvendor');
```

Define random variables

```
model.random; % random demand
model.random; % auxiliary random variable
model.ambiguity(S); % create ambiguity set
```

Define support sets for scenarios

```
for s = 1:S
    P(s).suppset(0 <= u, u <= ubar, norm(u - U(:,s)) <= v);
end
```

Define probabilities for scenarios

```
pr = P.prob;
pr.probset(pr == 1/S);
```

Define event-wise expectation

```
P.exptset(expect(u) <= theta);
```

Declare Wasserstein ambiguity set

```
model.with(P);
```

Define decision variables

```
w = model.decision(I,1);
y = model.decision(I,1);
```

Define event-wise adaptation

```
for s = 1:S
    y.evtadapt(s);
end
```

Define affine adaptation

```
y.affadapt(u);
y.affadapt(v);
```

Define objective function

```
model.min(-price'*w - expect(price'*y));
```

Define constraints

```
model.append(y >= 0);
model.append(y >= w - u);
model.append(w >= 0);
model.append(cost'*w == Gamma);
```

Solution

```
model.solve;
```
Experiment of Hanasusanto and Kuhn (2018)

We benchmark the RSO model against a state-of-art approximation scheme proposed by Hanasusanto and Kuhn (2018). Particularly, we repeat their experiment using the same set-ups.

Consider the second-stage problem of the form

\[ f(u) = \min\{ e^\top y \mid y \geq 0, y \geq Au - b \}, \]  

(30)

where \( e \) is a vector of ones. The problem does not have any here-and-now decision \( w \) and assumes that the random variable \( \tilde{u} \) resides in a box \( \mathcal{U} = [0, 1]^I_u \). Under the distance metric \( \rho(u, u^\dagger) = \|u - u^\dagger\|_2^2 \), Hanasusanto and Kuhn (2018) have shown that the worst-case expectation

\[ \sup_{P \in \mathcal{F}(\theta)} \mathbb{E}_P [f(\tilde{u})] \]  

(31)

amounts exactly to the optimal value of the following copositive program.

\[
\begin{align*}
\inf_{\alpha \in \mathbb{R}^S, \beta \in \mathbb{R}^+, \psi_s, \phi_s \in \mathbb{R}^{I_u + J_y}} & \frac{1}{2} \sum_{s \in [S]} (\alpha_s + \tilde{q}^\top \psi_s - \beta \|\tilde{u}_s\|_2^2 + \sum_{t \in [I_u + J_y]} \phi_{st} \tilde{q}_t^2) + \beta \theta^2 \\
\text{s.t.} & \left( \begin{array}{c}
\beta I + \tilde{Q}^\top \text{diag}(\phi_s) \tilde{Q} \\
-\frac{1}{2} T - \tilde{W} \text{diag}(\phi_s) \tilde{Q} \\
-\beta \tilde{u}_s^\top - \frac{1}{2} \psi_s \tilde{Q}
\end{array} \right) \tilde{\phi}_s \\
& \left( \begin{array}{c}
\frac{1}{2} (W \phi_s - \tilde{h})^\top
\end{array} \right) \geq_{K_{\text{cop}}} O \quad \forall s \in [S]
\end{align*}
\]

(32)

Here, \( K_{\text{cop}} \) is the copositive cone,

\[
\tilde{Q} = \begin{pmatrix} O \\ I \end{pmatrix}, \quad \tilde{q} = \begin{pmatrix} e \\ -e \end{pmatrix}, \quad T = \begin{pmatrix} A \\ O \end{pmatrix}, \quad \tilde{h} = \begin{pmatrix} -b \\ 0 \end{pmatrix}, \quad W = \begin{pmatrix} W & O \\ O & -I \end{pmatrix} \quad \text{with} \quad W = \begin{pmatrix} I \\ O \end{pmatrix},
\]

and \( O, I, 0 \) and \( e \), respectively correspond to the zero matrix, the identity matrix, the zero vector and the \( i \)-th standard unit basis, all of which are of the appropriate dimension. Because the copositive program (32) is generally intractable, Hanasusanto and Kuhn (2018) adopt a conservative \( K_0 \)-approximation by replacing the copositive cone \( K_{\text{cop}} \) with

\[ K_0 = \{ M \in S^K \mid M = P + N, P \succeq O, N \geq O \} \subseteq K_{\text{cop}}, \]

which leads problem (32) to a semidefinite program.

We run numerical tests for different pairs of the uncertainty dimension \( I_u \) and the sample size \( S \), and for each pair, we use the same set-ups as in Hanasusanto and Kuhn (2018) to generate 100 random instances. The Wasserstein radius is set to \( \theta = 1/S \). The dimension \( J_y \) of the recourse decision is sampled uniformly at random from \( \{1, 2, \ldots, \lceil \log(I_u + 1) \rceil \} \), \( A \) is sampled uniformly
Quite surprisingly, for all pairs of problem sizes, the solutions of both approximation approaches coincide for all 100 randomly generated instances. Unfortunately, we are not able to give a formal proof for this observation. Nevertheless, this observation supports that our proposed event-wise affine adaptation delivers solutions with competitive approximation quality as the state-of-the-art approximation scheme by Hanasusanto and Kuhn (2018). We report in Table 5 the average computation times of both approaches. In terms of computation efficiency, the event-wise affine adaptation outperforms because it leads to a second-order cone approximation to problem (32).

We note that the $K_0$-approximation by Hanasusanto and Kuhn (2018) also works when the cost vector of the second-stage problem (30) is affinely affected by the uncertainty, while our event-wise affine adaptation does not. On the other hand, the event-wise affine adaptation works with more general distance metrics and more general support sets that are not necessarily polyhedral (in the current experiment, the support set is a box $[0,1]^{I_u}$), while the $K_0$-approximation does not.

<table>
<thead>
<tr>
<th>$S$</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.3 &lt; 0.1</td>
<td>0.3 &lt; 0.1</td>
<td>0.3 &lt; 0.1</td>
<td>0.5 &lt; 0.1</td>
<td>2.3 &lt; 0.1</td>
</tr>
<tr>
<td>10</td>
<td>0.4 &lt; 0.1</td>
<td>0.4 &lt; 0.1</td>
<td>0.5 &lt; 0.1</td>
<td>0.9 &lt; 0.1</td>
<td>5.1 &lt; 0.1</td>
</tr>
<tr>
<td>20</td>
<td>0.6 &lt; 0.1</td>
<td>0.7 &lt; 0.1</td>
<td>0.8 &lt; 0.1</td>
<td>1.8 &lt; 0.1</td>
<td>11.6 &lt; 0.1</td>
</tr>
<tr>
<td>40</td>
<td>1.1 &lt; 0.1</td>
<td>1.3 &lt; 0.1</td>
<td>1.6 &lt; 0.1</td>
<td>3.5 &lt; 0.1</td>
<td>23.1 0.1</td>
</tr>
<tr>
<td>80</td>
<td>2.3 &lt; 0.1</td>
<td>2.5 &lt; 0.1</td>
<td>3.2 &lt; 0.1</td>
<td>7.8 &lt; 0.1</td>
<td>51.1 0.1</td>
</tr>
<tr>
<td>160</td>
<td>4.5 &lt; 0.1</td>
<td>5.1 &lt; 0.1</td>
<td>7.0 &lt; 0.1</td>
<td>18.0 &lt; 0.2</td>
<td>118.0 0.3</td>
</tr>
<tr>
<td>320</td>
<td>9.2 0.1</td>
<td>10.8 0.1</td>
<td>15.5 0.2</td>
<td>45.4 0.3</td>
<td>281.5 0.6</td>
</tr>
<tr>
<td>640</td>
<td>19.7 0.1</td>
<td>26.9 0.2</td>
<td>43.9 0.3</td>
<td>141.5 1.0</td>
<td>684.3 2.3</td>
</tr>
</tbody>
</table>

Table 5 Computation times (in seconds) of $K_0$-approximation (left) and event-wise affine adaptation (right).
D. Sample Code for K-means Adaptive Rebalancing

We assume \( E_k = \{ (k-1)K_2+1, \ldots , kK_2 \} \subseteq [S] \) for all \( k \in [K_1] \). Correspondingly, \( \kappa(s) = \left\lceil \frac{s}{K_2} \right\rceil \) for all \( s \in [S] \). We take a convex function \( \phi \) that specifies the mean absolute deviation of each random return within a particular cluster. Hence for each \( s \in [S] \), the cluster-wise support set is given by

\[
Z_s = \{ (u^1, u^2, v^1, v^2) | D^1_{\kappa(s)} u^1 \leq f^1_{\kappa(s)}, D^2_s u^2 \leq f^2_s, v^1 \geq |u^1 - \hat{\mu}^1_{\kappa(s)}|, v^2 \geq |u^2 - \hat{\mu}^2_s| \},
\]

where each cluster is in fact a polyhedron and where \(| \cdot |\) applies component-wise. The estimates \( \{ D^1_k \}_{k \in [K_1]}, \{ f^1_k \}_{k \in [K_1]}, \{ \hat{\mu}^1_k \}_{k \in [K_1]}, \{ \hat{\sigma}^1_k \}_{k \in [K_1]} \) are contained in MATLAB cells \( D1, f1, mu1, sigma1 \) and similarly, \( \{ D^2_s \}_{s \in [S]}, \{ f^2_s \}_{s \in [S]}, \{ \hat{\mu}^2_s \}_{s \in [S]}, \{ \hat{\sigma}^2_s \}_{s \in [S]} \) are contained in \( D2, f2, mu2, sigma2 \).

```matlab
% I: number of stocks
% K1: number of first-layer clusters
% K2: number of second-layer clusters
% ps: probabilities of clusters
% epsilon: risk threshold
% a,b,d,eta: parameters

% Create RSOME model
model = rsome('portfolio');

% Define random variables
u = model.random(I,2); % random demand
v = model.random(I,2); % auxiliary random variable
P = model.ambiguity(K1*K2); % create ambiguity set

% Define support sets for scenarios
for s = 1:K1*K2
    P(s).suppset(D1{ceil(s/K2)}*u(:,1) <= f1{ceil(s/K2)}, ...
                  D2{s}*u(:,2) <= f2{s}, ...
                  v(:,1) >= abs(u(:,1) - mu1{ceil(s/K2)}), ...
                  v(:,2) >= abs(u(:,2) - mu2{s}));
end

% Define probabilities for scenarios
pr = P.prob;
P.probset(pr == ps);

% Define event-wise expectation
for k = 1:K1
    P((k-1)*K2+1:k*K2).exptset(expect(u(:,1)) == mu1{k}, ...
                                expect(v(:,1)) <= sigma1{k});
end
for s = 1:K1*K2
    P(s).exptset(expect(u(:,2)) == mu2{s}, ...
                  expect(v(:,2)) <= sigma2{s});
end

% Declare K-means ambiguity set
```
model.with(P);

% Define decision variables
w = model.decision(I,1);
x = model.decision(I,1);
xbar = model.decision(I,1);
y = model.decision;
delta = model.decision;

% Define event-wise adaptation
for k = 1:K1
    x.evtadapt((k-1)*K2+1:k*K2);
    xbar.evtadapt((k-1)*K2+1:k*K2);
end
for s = 1:K1*K2
    y.evtadapt(s);
end

% Define affine adaptation
y.affadapt(u);
y.affadapt(v);

% Define objective function
model.min(delta + expect((1/epsilon)*y));

% Define constraints
model.append(y >= 0);
model.append(y >= x'*diag(a)*(u(:,1)-u(:,2)) ...  
            - w'*diag(a)*u(:,1) + b*(w + xbar) - delta);
model.append(xbar >= abs(x - w));
model.append(b'*xbar <= eta);
model.append(x >= 0);
model.append(a'*w == d);
model.append(w >= 0);

% Solution
model.solve;

E. Endnote
All mathematical programs in numerical experiments are solved using MOSEK on an Intel Core (TM) @ 3.40 GHz with 8GB RAM. The semidefinite program related to the $K_0$-approximation is implemented using the CVX interface (Grant and Boyd 2008), while remaining models are implemented using our developed algebraic modeling package RSOME (available at https://sites.google.com/view/rsome).